An Approach to Non-Linear Bayesian Forecasting Problems with Applications.

by

Helio dos Santos MIGON

Thesis submitted to the University of Warwick for the degree of Doctor of Philosophy, April 1984.

CONTENTS

Abstract

This thesis is devoted to the analysis and modelling of time series and it is concentrated on models and techniques which are of practical value. In particular we developed a wide class of non-linear dynamic models which are useful in the handling of real life problems.

Initially we review the basic principles of Bayesian forecasting and the design of Dynamic Linear Models. The main body of the thesis attacks the problem of Normal non-linear estimation and forecasting. Some applications to the seasonal multiplicative model are exhaustively discussed. Following this we present the results of an application of Bayesian transfer response in Market Research. This application worked as the very first stimulus to extend the non-linear models to the exponential family.

Finally we discuss the concepts of stochastic transfer response modelling and associated sequential estimation, and we report some applications of the method and models for long term forecasting.

Acknowledgements.

I should like to express my sincere gratitude to Professor P.J. Harrison for his efforts in initiating this research, his continuous suggestions, guidance and general support throughout the period of study.

I am also grateful to Dr. M. West for his discussions and comments and to Dr. P. Walley for his kindness and attention.

Thanks are also due to Jean Lake, not only for her excellent typing skills, but for her patience and cheerfulness.

I am deeply grateful to Mirna and the three Musketeers for their patience, endurance, encouragement and many sacrifices during my period of study.

This research was funded by Conselho Nacional de Desenvolvimento Científico & Tecnológico - CNPq - and I should like to thank the financial support of the Universidado Federal do Rio de Janeiro - UFRJ.

To

Mirna

Marcio

Marcelo

Renato

Chapter 1: Introduction.

1.1 General.

This thesis is concerned with the analysis and modelling of time series and it is concentrated on models and techniques which are of practical value. It was our decision to develop new models but emphasize the applied and methodological aspects of the problem rather than to look for theoretical results.

Time series models can be interpreted as models which are constructed without drawing on any theories concerning possible behavioural relationship between variables. This is a distinguishing feature of a time series, as opposed, say to an econometric model. This field has received an enormous amount of interest from workers in socio-economics studies, physical and engineering sciences and in areas such as demography and medicine.

The classical models for time series have been developed since the 40's and they are based on ARMA processes. A full strategy for time series analysis was developed by Box and Jenkins in the 70's.

It is worth pointing out that they have started with results of Kolmogorov and Wiener and are based on the hypothesis of derivable stationarity with predictors as linear functions of the past observations and mean square error as criterion of optimality.

The works of Harrison and Stevens (1971, 1976) on Bayesian Forecasting overcome some of the restrictive hypotheses and opened a very new area for: practical and theoretical investigation. The approach gives a logically consistent description of the way in which a forecasting method should deal with a wide class of situations.

The principle of superposition enables an easy way of building linear models from separate components, which are structured and facilitates the meaning.

The Bayesian approach to statistics, although still somewhat controversial, provides a rich and logical framework for mathematical modelling. This is, for example, the natural way to blend subjective information with data in order to produce inference. Furthermore, the combination of prior and experimental information is rigorous and natural.

The Dynamic Linear Models and the Bayesian Forecasting of Harrison and Stevens have been extended by Ameen and Harrison (1983). The introduction of a class of Normal Discount Bayesian Models, founded upon the discount concept, has overcome some difficulties found by practitioners. The innovation or variance matrix is replaced by a set of discount factors which are invariant to the measurement scale of the independent and the control variables. This introduces conceptual simplicity and also removes much of the ambiguity. The flexibility of these models and their potential as aid to understanding physical systems as well as forecasting, suggests that more interest will be centred on their application to diverse fields in the near future.

This thesis is devoted to the examination of a wide class of non-linear models which, in particular, extends the work of Nelder and Wedderburn (1972) in Generalized Linear Models. The appropriate sampling distribution should be used as in West, Harrison and Migon (1983). Some direct application of the non-linear dynamic models to the estimation of stochastic response is also discussed.

1.2 Outline of the thesis.

Chapter 2 is concerned with a review of Dynamic Linear Model (D.L.M.) design and the relationship between constant DLM's and ARIMA models. The main point is concerned with the design of canonical models which produce desired forecasting functions. These canonical models can be transformed using similar linear operators in order to produce a convenient meaningful parametrization.

In Chapter 3 we present a class of non-linear normal dynamic models which provides a much wider class of applications of the dynamic models to real life problems. This chapter begins with a review of the normal weighted Bayesian model and reformulates it in order to put forward the non-linear model. Finally, we present an application to the linear growth seasonal model in a multiplicative form which avoids transformations and so relates the parameters in the original measurement scale which is preferred by practitioners.

Chapter 4 consists of a report on an application of Bayesian transfer response modelling to a marketing problem. At the very beginning of our study on transfer response, an opportunity to participate in a project involving the Statistics Department at Warwick University and Millward-Brown Marketing Research Co. arose. The models developed have been working successfully for more than two years in their computers and an early practical evaluation can be found in Colman and Brown (1983). Altogether we developed four models, being two linear and two non-linear. The basic linear model is a dynamic extention of Broadbent's model (Broadbent (1979)). This fortunate application acted as the very first stimulus for the methodological developments presented in this thesis.

An extention of the Normal non-linear models is presented in Chapter 5. This model can be viewed as a dynamic version of the generalized linear model of Nelder and Wedderburn (1972).

The linear Bayes approach developed by Hartigan (1969) is extensively used. This method is appropriate to problems of inference when only the first two moments of the prior distribution and the likelihood are specified. Some applications for the log-normal and binomial cases are discussed and for additional applications the paper of West,Harrison and Migon (1983) is recommended.

Chapter 6 considers the on-line estimation of transfer response. The concept of stochastic transfer response is discussed and some parallels drawn with the classical approach. The performance of the estimation method was assessed by a simulation experiment.

Finally, in Chapter 7, we present an application of the non-linear Bayesian forecasting method to the long-term forecasting problem. The low frequency or trend is represented by a growth curve such as the modified exponential and the high frequency as an autoregressive process of order p. The parameters are estimated on-line and it is worth pointing out that the principle of superposition is again used. This method was applied to some real data and the main results are reported.

The paper of Harrison and Akram (1982) is recommended for a theoretical discussion of the problem.

1.3 Terminology and Notation

In general we tried to keep the notation as close as possible to other related works. Throughout the thesis all probability distributions are defined via densities with respect to Lebesque measure, and they will be represented by the generic symbol $p(\cdot)$. Some special densities are denoted as:

(i) $N_{\underline{\theta}}(\underline{m},\underline{C})$ is the multivariate normal density of $\underline{\theta}$, with mean vector $\underline{m}$ and variance matrix $\underline{C}$;

(ii) $G_{\theta}(a,b)$ is the Gamma density for θ;

(iii) $Be(r,s)$ is the Beta density for θ; and

(iv) $B(n,p)$ is the Binomial mass probability function.

The general exponential family has a density given as:

$$p(Y/\psi,\phi) = \exp[\phi\{Y\psi+a(\psi)\}]\ b(y,\phi)$$

where $\phi>0$ and ψ is the natural parameter of the distribution. It includes as special cases the Normal, Gamma, Poisson and Binomial distributions.

No distinction is made between random variables and their observed value since, generally, the context will be clear. By $x \sim p(\cdot)$ we mean that x has density $p(\cdot)$. For example: $(\underline{x}/\underline{y}) \sim N[\underline{m},\underline{C}]$ represents the conditional multivariate normal density of $\underline{x}$ given $\underline{y}$; and, in general, $(\underline{x}/\underline{y}) \sim [\underline{m}, \underline{C}]$ means that the conditional density is partially specified with mean $\underline{m}$ and variance matrix $\underline{C}$.

All the vectors are underlined, as $\underline{x}$, for example; and matrices appear as capital letters.

The following abbreviations represent the classes of models discussed in this thesis.

DLM Dynamic Linear Models;

CDLM Constant Dynamic Linear Models;

NDLM Normal Dynamic Linear Models;

NDBM Normal Discount Bayesian Models;

NWBM Normal Weighted Bayesian Models;

ARIMA Autoregressive integrated moving average models; and

GLM Generalized Linear Models

Finally, it is worth remarking that the equations are numbered according to the chapters and sections. For example: equation number 4.3.2 means equation number two in section 3 of Chapter 4. Further notation will be introduced as necessary.

Chapter 2: A Review of DLM and ARIMA Models.

2.1 Introduction

This chapter is concerned with a review of the Dynamic Linear Model, model design and the relationship between the class of constant DLM and ARIMA models.

In Section 2.2 the Bayesian forecasting approach of Harrison-Stevens (1971 & 1976) is discussed. A more suitable Bayesian formulation is presented. The concept of similar models which is very useful in model design is introduced in Section 2.3, as well as a strategy to DLM modelling.

Finally in Section 2.4 we show the link between CDLM and ARIMA models. The Box-Jenkins strategy is briefly discussed and a state space representation is introduced.

2.2 Normal Dynamic Linear Model

The Bayesian forecasting approach developed by Harrison & Stevens (1976), a reference which we shall abbreviate HS(1976), is based on a complete parametric description of the process, which is incorporated into a dynamic linear set of equations describing:

(i) the process observation, and

(ii) the parameter evolution

The parametric structure reduces ambiguity in the model building and so increases the meaning, which permits communication between the decision maker, the forecaster and the forecast method in both directions. When building larger models the forecaster is almost forced into model decomposition, and again the structure has a great deal to offer.

The dynamics of the system is described by the evolution of the parameters in time, both as a result of the inherent process and from random shocks as disturbances.

In its general form the Dynamic Linear Model is stated in terms of an ordered index which may be regarded as discrete points of time. Such a model is defined by a quadruple $\{F,G,V,W\}_t$ for each time index. The model may then be written:

$$\text{observation equation:} \quad \underline{Y}_t = \underline{F}_t\, \underline{\theta}_t + \underline{\nu}_t \quad ; \quad \underline{\nu}_t \sim N[\underline{O}, \underline{V}_t]$$

$$\text{system equation:} \quad \underline{\theta}_t = \underline{G}\, \underline{\theta}_{t-1} + \underline{w}_t \quad ; \quad \underline{w}_t \sim N[\underline{O}, \underline{W}_t] \qquad (2.2.1)$$

where: $\underline{Y}_t$ is an m×1 vector of observations at time t;

$\underline{\theta}_t$ is an p×1 vector of parameters at time t;

$\underline{F}_t$ is an m×p matrix of independent variables, known at time t;

$\underline{G}$ is a known p×p matrix defining the parameter evolution;

$\underline{\nu}_t$ is an m×1 vector representing the observation noise;

$\underline{w}_t$ is an p×1 vector representing the parameter noise.

In general $\underline{\nu}_t$ is assumed $N(\underline{O},\underline{V}_t)$ and $\underline{w}_t$ is assumed $N(\underline{O}, \underline{W}_t)$ with each being a series of independent random vectors. We should emphasize that the DLM can be generalized by allowing $\underline{G}$ to vary throughout the time and by considering a non zero expected value for $\underline{\nu}_t$ and $\underline{\omega}_t$.

An alternative formulation.

Equation 2.2.1 can be rewritten in a more appropriate Bayesian statement for the DLM $\{F, G, V, W\}_t$, as:

$$\begin{cases} \text{observation equation:} & (\underline{Y}_t|\theta_t) \sim N[\underline{F}_t\, \underline{\theta}_t; \underline{V}_t] \\ \text{parameter relation:} & (\underline{\theta}_t|\underline{\theta}_{t-1}) \sim N[\underline{G}\, \underline{\theta}_{t-1}, \underline{W}_t] \end{cases} \qquad (2.2.2)$$

Information

Definition 2.1: D_{t-1} represents all the available information at t-1.
This includes all available data and subjective information.

However, unless otherwise stated $D_t = \{D_{t-1}, y_t, \underline{F}_t\}$.

At time t-1 assume that the posterior parameter information is

$$(\underline{\theta}_{t-1}|D_{t-1}) \sim N[\underline{m}_{t-1}, \underline{C}_{t-1}];$$

it then follows that given the DLM a future joint forecast distribution can be obtained for any required time. Further, on reception of the actual observation, Bayes Theorem can be applied to obtain the posterior parameter distribution $(\underline{\theta}_t|D_t)$.

Updating and Forecasting.

The Markovian nature of the model lends itself naturally to a sequential estimation of $\underline{\theta}_t$, forecasting future $\underline{Y}_{t+k}$, $k\geq 1$, and smoothing or filtering, i.e. "forecasting" into the past values of $\underline{\theta}_{t+k}$, $k = -1, -2...$

The two major operations in the sequential analysis are:

(i) Time update

$$p(\underline{\theta}_t|D_{t-1}) = \int p(\underline{\theta}_t|\underline{\theta}_{t-1})\ p(\underline{\theta}_{t-1}|D_{t-1})\ d\underline{\theta}_{t-1}$$

(ii) Prior to posterior update

$$p(\underline{\theta}_t|D_t) \propto p(\underline{\theta}_t|D_{t-1})\ p(y_t|\underline{\theta}_t, D_{t-1})$$

where D_{t-1} is as in definition 2.1 and represents all the information up to and including time t-1.

If $p(\underline{\theta}_t|D_{t-1})$ is a $N(\underline{G}\ \underline{m}_{t-1}, \underline{R}_t)$ then once we observe $\underline{Y}_t = \underline{y}_t$, the parameter posterior distribution at time t is obtained by applying Bayes Theorem:-

$$\lg\ p(\underline{\theta}_t|D_{t-1}y_t) = \lg\ L(\underline{y}_t|\underline{\theta}_t) + \lg\ p(\underline{\theta}_t|D_{t-1}) + \text{const, where}$$

$L(\underline{y}_t \mid \underline{\theta}_t)$ is the likelihood function. This immediately gives $(\underline{\theta}_t|D_t) \sim N(\underline{m}_t, \underline{C}_t)$ with the following recurrence relationship for $\underline{m}_t$ and $\underline{C}_t$:

$$\underline{C}_t^{-1} = \underline{R}_t^{-1} + \underline{F}_t'\ \underline{V}_t\ \underline{F}_t$$

$$\underline{m}_t = G\ \underline{m}_{t-1} + \underline{A}_t\ \underline{e}_t \qquad (2.2.3)$$

where as usual in Bayesian Normal Theory $e_t = \underline{Y}_t - \underline{F}_t\ \underline{G}\ \underline{m}_{t-1}$ is the one step ahead forecast error and $\underline{A}_t$ is the multiple regression vector of $\underline{\theta}_t$ on $\underline{Y}_t$, often called the gain vector in control engineering. Alternative forms for the precision recurrence in terms of variances are:

$$\underline{C}_t = (\underline{I}-\underline{A}_t\ \underline{F}_t)\ \underline{R}_t \tag{2.2.4}$$

or

$$\underline{C}_t = \underline{R}_t - \underline{A}_t\ \hat{\underline{Y}}_t\ \underline{A}'_t, \quad \text{where}$$

$\hat{\underline{Y}}_t = \text{var}(Y_t|D_{t-1})$. The latter variance equation together with equation for $\underline{m}_t$ are often referred to as the Kalman filter recurrence equations [H.S.(1976)].

Forecasting distribution.

The predictive distribution is easily obtained, in an extrapolative way, as soon as the posterior distribution of the parameter vector $\underline{\theta}_t$ at any time $t\geq 1$ is evaluated. The predictions are distributional in nature and derived from the current parameters uncertainty, future observation noise $\underline{V}_{t+k}$ and the disturbances $\underline{W}_{t+k}$, k=1,2...

From the DLM equation (2.2.1), a future observation variable is written as:

$$\begin{aligned} \underline{Y}_{t+k} &= \underline{F}_{t+k}\ \underline{\theta}_{t+k} + \underline{\nu}_{t+k} \\ \underline{\theta}_{t+k} &= \underline{G}\ \underline{\theta}_{t+k-1} + \underline{w}_{t+k} \end{aligned} \tag{2.2.5}$$

Defining

$$\begin{aligned} \hat{\underline{m}}_{k,t} &= E(\underline{\theta}_{t+k}|D_t, \underline{F}_{t+k}) \\ \hat{\underline{C}}_{k,t} &= \text{var}(\underline{\theta}_{t+k}|D_t, \underline{F}_{t+k}) \end{aligned} \tag{2.2.6}$$

where: $\underline{m}_{0,t} = \underline{m}_t$

$\underline{C}_{0,t} = \underline{C}_t$ are known from the recurrence equations (2.2.3).

Then from (2.2.5) we obtain

$$\hat{\underline{m}}_{k,t} = \underline{G}\,\hat{\underline{m}}_{k-1,t}$$

$$\hat{\underline{C}}_{k,t} = \underline{G}\,\hat{\underline{C}}_{k-1,t}\,\underline{G}' + \underline{W}_{t+k} \qquad (2.2.7)$$

with $\underline{V}_{t+k}$ and $\underline{W}_{t+k}$ known.

The prediction mean and variance of the k-steps-ahead process are written as

$$\hat{\underline{y}}_{k,t} = E(\underline{Y}_{t+k}|D_t, \underline{F}_{t+k})$$

$$\hat{\underline{Y}}_{t+k} = \mathrm{var}(\underline{Y}_{t+k}|D_t, \underline{F}_{t+k}) \qquad (2.2.8)$$

and two cases must be considered in order to calculate these values.

Case I: $\underline{F}_{t+k}$, k=1,2... known

From (2.2.5) and (2.2.6) we get:

$$\hat{\underline{y}}_{k,t} = \underline{F}_{t+k}\,\hat{\underline{m}}_{k,t}$$

$$\hat{Y}_{k,t} = \underline{F}_{t+k}\,\hat{\underline{C}}_{k,t}\,\underline{F}'_{t+k} + V_{t+k}$$

with $\hat{\underline{m}}$ and $\hat{\underline{C}}$ obtained recursively for k=1,2... from (2.2.6).

Case II: $\underline{F}_{t+k}$, k=1,2... not known.

Let $\underline{F}_{t+k}$ be expressed by an expected value $\hat{\underline{F}}$ plus a stochastic term $\delta\underline{F}$, where for simplicity we drop the suffix t+k.

$$\underline{F}_{t+k} = \hat{\underline{F}} + \delta\underline{F}; \qquad \delta\underline{F} \sim N(0;\Phi)$$

Then using a result of Feldstein (1971) we get

$$\hat{\underline{y}}_{k,t} = \hat{\underline{F}}_{k,t} \hat{\underline{m}}_{k,t}$$

$$\hat{\underline{Y}}_{k,t} = \hat{\underline{F}}_{k,t} \hat{\underline{C}}_{k,t} \hat{\underline{F}}'_{k,t} + \underline{V}_{t+k} + \underline{U}$$

where: $\underline{U} = (u_{ij})$ is a p×p matrix with elements

$$u_{ij} = tr\{\Phi_{ij} [\hat{\underline{m}}_{k,t} \hat{\underline{m}}'_{k,t} + \hat{\underline{C}}_{k,t}]\}$$

Finally we would emphasize the basic characteristics of the DLM, some of which will be largely used in this thesis:

(i) its easy physical interpretation for the parameters, designed to the requirements of the decision makers;

(ii) its probabilistic information on the parameters at any time;

(iii) its sequential nature which permits a description of systematic changes in the parameters of a system;

(iv) its freedom from stationarity which allows a better description of "reality", and

(v) its distributional predictive nature.

To complete this section it should be mentioned some new theoretical and methodological developments in the area of Bayesian Forecasting as the Normal Discount Bayesian Model [Ameen and Harrison (1983)]; the Generalized Exponential Weight Regression [Harrison and Akram (1982)]; and the work in robustification of DLM [West (1981)], the non-normal steady state models of Smith (1979) and Souza (1981). In the applications of DLM and Bayesian forecasting the papers of Johnston and Harrison(1980) and Harrison et al (1977) are related with our thesis.

2.3 Model Design

2.3.1 Introduction and definitions.

The aims of this section are to show how to design models with a desired forecast function. The meaning of forecast function is the expected value of the forecast distribution and it is a local form description of the expected process trajectory. It is clear that the forecast function is completely determined by the evolution matrix $\underline{G}$, assuming $\underline{F}_{t+k}$, k=1,2,... known.

A general Normal Dynamic Linear Model is denoted by $\{\underline{F}, \underline{G}, \underline{V}, \underline{W}\}_t$, with parameterization $\underline{\theta}$, where $(\underline{\theta}_t|D_t) \sim N[\underline{m}_t, \underline{C}_t]$. Following the same notational principals a Constant Normal Dynamic Linear Model is $\{\underline{F}, \underline{G}, \underline{V}, \underline{W}\}$ with information $(\underline{\theta}_t|D_t) \sim N[\underline{m}_t, \underline{C}_t]$.

In order to characterize a class of similar models we have to introduce the concept of observability. Loosely a deterministic system is said observable if knowledge of y_t's over a finite period of time completely determine the parameter $\underline{\theta}_t$. Consequently for a constant D.L.M. one can think of unobservable model as an over-specified model.

Definition 2.1: The deterministic model $\{\underline{F}, \underline{G}, \underline{0}, \underline{0}\}$ is said to be observable if and only if $\underline{T} = (\underline{F}, \underline{F}\,\underline{G}, \ldots \underline{F}\,\underline{G}^{p-1})'$ is a p×p matrix of full rank p.

The definition 2.1 is extended for a constant Normal DLM $\{\underline{F}, \underline{G}, \underline{V}, \underline{W}\}$ in a natural way saying that the constant NDLM is observable if the associated deterministic model is observable.

2.3.2. Similar Models and Reparametrization.

In practice it is very important to transform models in order to get a meaningful parameterization, so that forecast and information are easily understood, and subjective knowledge can be incorporated smoothly.

Definition 2.2: Two observable constant NDLM, say

$$M \equiv \{\underline{F}, \underline{G}, \underline{V}, \underline{W}\} \text{ and } M_1 \equiv \{\underline{F}_1, \underline{G}_1, \underline{V}, \underline{W}_1\}$$

are defined as similar models if and only if:

(i) $\underline{G}$ is similar to $\underline{G}_1$, i.e. there exist a full matrix H such that $\underline{G}_1 = \underline{H}\,\underline{G}\,\underline{H}^{-1}$

(ii) $\underline{W}_1 = \underline{H}\,\underline{W}\,\underline{H}^{-1}$, where $\underline{H}$ is a full rank matrix such that $\underline{H}\,\underline{G}\,\underline{H}^{-1} = \underline{G}_1$, and

(iii) $\underline{F}_1 = \underline{F}\,\underline{H}^{-1}$

Theorem 2.3: Let $\underline{\theta}$ be the parameterization of the model M and $\underline{\psi}$ be that of model M_1. Suppose that $(\underline{\theta}_t|D_t) \sim N[\underline{m}_t, \underline{C}_t]$ and $(\underline{\psi}_t|D_t) \sim N[\underline{m}_{1,t}, \underline{C}_{1,t}]$ and that M and M_1 are similar observable models. Then these models have the same forecasting function if:

$$\underline{m}_{1,t} = \underline{H}\,\underline{m}_t;\ \text{with}\ \ \underline{H} = \underline{T}_1^{-1}\ \underline{T},\ T^{-1},\ \text{exist because } M_1 \text{ is observable}$$

where $\underline{T} = (\underline{F}, \underline{F}\,\underline{G},\ldots,\underline{F}\,\underline{G}^{p-1})'$ and $\underline{T}_1 = (\underline{F}_1,\underline{F}_1\,\underline{G}_1,\ldots,\underline{F}_1\,\underline{G}_1^{p-1})'$

Proof: (i) $\underline{F}_1\,\underline{G}_1^k\,\underline{m}_{1,t} = \underline{F}\,\underline{H}^{-1}[\underline{H}\,\underline{G}^k\,\underline{H}^{-1}]\,H\,\underline{m}_t = \underline{F}\,\underline{G}^k\,\underline{m}_t$, $\forall k \geq 0$

(ii) Because M_1 is observable the inverse of $\underline{T}_1$ exists, and from (i) we can write: $\underline{T}\,\underline{m}_t = \underline{T}_1\,\underline{m}_{1,t}$. Then follows that $H = \underline{T}_1^{-1}\,\underline{T}$

To complete this section we emphasize that having a model M with parameterization $\underline{\theta}$ we can reparametrize it through the transformation $\underline{H}$: $\underline{\theta} \mapsto \underline{\psi}$, obtaining the similar model $M_1 \equiv \{\underline{F}\,\underline{H}^{-1}, \underline{H}\,\underline{G}\,\underline{H}^{-1}, \underline{V}, \underline{H}\,\underline{W}\,\underline{H}^{-1}\}$ with parametrization $\underline{\psi}$, as stated in Ameen and Harrison (1983).

2.3.3 Canonical form models.

The design of a DLM with some given form to the forecast function, i.e. the trajectory of the process, can be done in two steps:

(i) write down the canonical design with parametrization $\underline{\psi}$, and

(ii) reparametrize to any required parametrization, $\underline{\theta} = \underline{H}\,\underline{\psi}$, as discussed in section 2.3.2.

For the model $\{\underline{F}, \underline{G}, \underline{V}, \underline{W}\}$ with parametrization $(\underline{\theta}_t|D_t) \sim N(\underline{m}_t, \underline{C}_t)$ assuming $\underline{F}_{t+k}$ known, the forecast function

$$F_t(k) = E(Y_{t+k}|D_t)$$
$$= \underline{F}_{t+k}\,\underline{G}^k\,\underline{m}_t$$

is completely described by the system matrix $\underline{G}$.

Using the well known Jordan Decomposition theorem we find a non-singular matrix $\underline{H}$, such that

$$\underline{H}\,\underline{G}\,\underline{H}^{-1} = \Lambda = \mathrm{diag}(\lambda_1 \ldots \lambda_p) \text{ if } \underline{G}$$

has p distinctive eigenvalues $\lambda_1 \ldots \lambda_p$, and the model M is similar to the model $M_1 \equiv \{\underline{F}_1\ \underline{G}_1\ \underline{V}\ \underline{W}_1\}$. Hence these models have the same forecast function.

It is not difficult to show that:

$$\begin{aligned} F_t(k) &= [\underline{F}\,\underline{H}^{-1}]\,\underline{H}\,\underline{G}^k\,\underline{H}^{-1}\,[\underline{H}\,\underline{m}_t] \\ &= \underline{F}\,\underline{H}^{-1}\,\Lambda^k\,\underline{H}\,\underline{m}_t \\ &= \sum_{i=1}^{p} a_{i,t}\,\lambda_i^k \text{ , with } a_{i,t} = m_{i,t}. \end{aligned}$$

Now if $\underline{G}$ is of rank p and has r<p distinctive eigenvalues $\lambda_1 \ldots \lambda_p$ each one with multiplicity p_i, i=1,...,r, with $\sum_{i=1}^{p} p_i = p$, there exist $\underline{H}$ non-singular such that

$$\underline{H}\,\underline{G}\,\underline{H}^{-1} = \mathrm{diag}\big(\Lambda_1(\lambda_1)\ldots\Lambda_r(\lambda_r)\big), \text{ where:}$$

$$\Lambda_i(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & 0\ldots0 \\ 0 & \lambda_i & 1\ldots0 \\ \vdots & & \vdots \\ 0 & 0 & 0\ldots\lambda_i \end{bmatrix}$$ is a Jordan block of dimension p_i.

In this case the forecast function is:

$$\begin{aligned} F_t(k) &= [\underline{F}\,\underline{H}^{-1}]\,\underline{H}\,\underline{G}^k\,\underline{H}^{-1}\,[\underline{H}\,\underline{m}_t] \\ &= \sum_{i=1}^{r} \left[\sum_{j=1}^{p_i-1} a_{i,j,t}\,k^j\right] \lambda_i^k \\ &= \sum_{i=0}^{r} P_{i,t}(k)\,\lambda_i^k \text{ , where } P_{i,t} \end{aligned}$$

is a polynomial of order p_i in k, with degree p_i-1.

Example: A simple cycle is characterized by a pair of distinct eigenvalues (e^{iw}, e^{-iw}). The model in canonical form is completely undesirable since the parameters and the random disturbances are complex random variables. Denoting the model in canonical form as $M \equiv \{(1,1); \begin{bmatrix} e^{iw} & 0 \\ 0 & e^{-iw} \end{bmatrix}; V, \underline{W}\}$ with parametrization $\underline{\theta}$, where $(\underline{\theta}_t | D_t) \sim N[\underline{m}_t, \underline{C}_t]$ follows that the forecasting function is:

$$F_t(k) = m\, e^{iwk} + \overline{m}\, e^{-iwk},$$

where: $m = \frac{a}{2} + i\frac{b}{2}$ and $\overline{m} = \frac{a}{2} - i\,\frac{b}{2}$.

The forecasting function can be rewritten as:

$$F_t(k) = a \cos(kw) - b \sin(kw)$$

and it motivates the following model:

$$M_1 \equiv \left\{(1,0); \begin{bmatrix} \cos(w) & \sin(w) \\ -\sin(w) & \cos(w) \end{bmatrix}; V, \underline{W}_1\right\}.$$

Of course these models are similar and the transformation is defined as:

$\underline{H} = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} = \underline{T}_1^{-1}\, \underline{T}$. The new parametrization is

$$\underline{\psi}_t = \underline{H}\, \underline{\theta}_t = \begin{pmatrix} \theta_{1,t} + \theta_{2,t} \\ \theta_{1,t} - i\,\theta_{2,t} \end{pmatrix} \quad \text{and}$$

$$(\underline{\psi}_t | D_t) \sim N\left[\begin{bmatrix} a \\ b \end{bmatrix}; \underline{C}_{1,t}\right], \text{ where } \underline{C}_{1,t} = \underline{H}\, \underline{C}_t\, \underline{H}^{-1}$$

Finally it is worth showing that the forecast function for the model $M \equiv \{\underline{F}, \underline{G}, \underline{V}, \underline{W}\}$ can be written as a linear function of $\underline{S}_t = \left(F_t(0) \ldots F_t(p-1)\right)'$

Let $\underline{T} = (\underline{F}, \underline{F}\,\underline{G} \ldots \underline{F}\,\underline{G}^{p-1})'$ be of full rank p (the model is observable). Then $F_t(0) \ldots F_t(p-1)$ are linearly independent and $\forall k > 0$ exist $\underline{\ell}_k = (\ell_0 \ldots \ell_{p-1})_k \in \mathbb{R}^p$ such that:

$$F_t(k) = \underline{\ell}_k\, \underline{S}_t = \sum_{i=0}^{p-1} \ell_{i,k}\, F_t(i)$$

Proof: $\underline{S}_t = \underline{T}\, \underline{m}_t$ implies $\underline{m}_t = \underline{T}^{-1}\, \underline{S}_t$.

Hence $F_t(k) = \underline{F}\, \underline{G}^k\, T^{-1}\, \underline{S}_t = \underline{\ell}_k\, \underline{S}_t$

$$= \sum_{i=0}^{p-1} \ell_{i,k}\, F_t(i)$$

Example: $M = \{(1,0); \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}; V, \underline{W}\}$ and $(\underline{\theta}_t | D_t) \sim N(\underline{m}_t; \underline{C}_t)$. For this example

$\underline{T} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, $\underline{G}^k = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$ and $S_t = \begin{pmatrix} F_t(0) \\ F_t(1) \end{pmatrix}$, $\underline{T}^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$. Hence

$$F_t(k) = \underline{F}\ \underline{G}^k\ \underline{T}^{-1}\ \underline{S}_t$$

$$= F_t(0) + k\left[F_t(1) - F_t(0)\right]$$

$$= m_{1,t} + k\ m_{2,t} \quad \text{as is well known.}$$

2.4 Constant DLM and ARIMA Models.

2.4.1 ARIMA Models.

Since the developments of Wiener and Kolmogorov in 1940's many studies have been done based in the hypothesis of stationarity or derivable stationarity, with predictors as linear functions of the past observations and mean square error as criterion of optimality.

Box and Jenkins in a series of papers and a book (1976) describe a full strategy for the construction of linear stochastic difference equations describing the behaviour of a time series. Briefly, they assume that the given series Y_t can be reduced to stationarity by differencing a finite number of times, i.e. by determining the stationary series Z_t as:

$$Z_t = (1-B)^d\ Y_t \quad \text{with } d>0 \text{ integer}$$

and B the backward shift operator in the time index of Y_t, i.e. $B\ Y_t = Y_{t-1}$

An ARMA model is assumed for the stationary time series Z_t and it is written as:

$$(1-\phi_1 B-\phi_2 B^2 \ldots -\phi_p B^p)Z_t = (\theta_0-\theta_1 B-\ldots-\theta_q B^q)a_t, \qquad (2.4.1)$$

where: a_t, t=1,2... is a sequence of uncorrelated random variables with $E[a_t] = 0$ and $\gamma_a(B) = \sigma^2$ is the auto covariance generating function;

ϕ_i, i=1,2...p are the autoregressive parameters, and

θ_i, i=1...q are the moving average parameters.

In terms of the original time series we can write:

$$\Phi(B)(1-B)^d Y_t = \Theta(B)a_t \tag{2.4.2}$$

where $\Phi(B)$ and $\Theta(B)$ are polynomials in B of degree p and q respectively. These models are the well known ARIMA(p,d,q).

It is worth mentioning the Box-Jenkins procedure to fit a model of the above form (2.4.2) to a given set of data. It consists of a three steps procedure:

(i) identification of the order of the model;

(ii) estimation of parameters, i.e. ϕ's, θ's and σ^2, and

(iii) a diagnostic check.

These three steps are used iteratively and the popularity of the method is **mainly** based in the large quantity of software and relevant literature available.

It is straightforward to show that the mean square error forecast of Y_{t+k} from 2.4.2 is its conditional expectation. So rewriting 2.4.2 as $\Phi(B)Y_{t+k}=\Theta(B)a_t$ and taking conditional expectation noting that the conditional expectation of Y_t, $Y_{t-1},\ldots$, a_t, $a_{t-1}\ldots$ are their known values at time t, and those for a_{t+k} are zero, we get the forecasting as a step by step procedure, taking in turn k=1,2,...

2.4.2 The state space representation.

There is a simple and interesting way of obtaining a state space or Markovian representation for the models described in sec. 2.4.1. The basic ideas rest on the well known result that any finite order difference or differential equation can be rewritten as a vector first order equation. Thus we may rewrite 2.4.2 in an algebraically equivalent form:

$$\begin{aligned} Y_t &= \underline{F}\,\underline{\theta}_t \\ \underline{\theta}_t &= \underline{G}\,\underline{\theta}_{t-1} + \underline{\Gamma}\,u_t \end{aligned} \tag{2.4.3}$$

where: $\underline{\theta}_t$ is a vector (n+1)×1 called the state space vector and $\underline{F}$, $\underline{G}$ and $\underline{\Gamma}$ are suitably defined matrices.

It should be emphasized that the specific expressions for $\underline{F}$, $\underline{G}$ and $\underline{\Gamma}$ provide merely one way of writing the state space form. Even if we fix the dimension of $\underline{\theta}$, we may still obtain an equivalent state space representation by making a similarity transformation as described in section 2.3.2.

Following Harrison and Akram (1983) a canonical state space representation for the model:

$$\Phi(B)\, Y_t = \theta(B)\, u_t \qquad \text{, where}$$

$\Phi(B)$ and $\theta(B)$ are polynomial in B of degree p and q respectively, can be constructed using:

$$\underline{F} = (1,0\ldots0) \quad ; \underline{G} = \begin{bmatrix} -\underline{\alpha} & I_n \\ 0 & \underline{0} \end{bmatrix} ;$$

$\underline{\alpha} = (\phi_1,\ldots\phi_k)'$ and $\underline{\Gamma} = (\theta_0\theta_1\ldots\theta_n)$, where $n = \max(p,q)$ and $\phi_i=0$ if $i>p$, $\theta_j=0$ if $j>q$.

2.4. The Relationship between CDLM and ARIMA Models.

To complete this chapter we show that any closed Dynamic Linear Model can be represented as an ARIMA model. The proof of the theorem 2.7 follows from a straightforward application of the following lemmas.

Lemma 2.5: If in the deterministic CDLM, i.e. $\{\underline{F}, \underline{G}, 0, 0\}$, the matrix $\underline{G}$ has eigenvalues $\lambda_1\ldots\lambda_p$ not necessarily distinct nor necessarily non zero then

$$\prod_{i=1}^{p} (1-\lambda_i B)\, Y_t = 0$$

Proof: $\qquad (1-\lambda_i B)Y_t = \underline{F}[\underline{G}-\lambda_i I]\underline{\theta}_{t-1}$ and hence

$$\prod_{i=1}^{p} (1-\lambda_i B)Y_t = \underline{F}\left[\prod_{i=1}^{p} (\underline{G}-\lambda_i I)\right]\underline{\theta}_{t-p} \qquad \text{but}$$

$f(x) = \prod_{i=1}^{p} (x-\lambda_i) = 0$ is the characteristic equation of $\underline{G}$ and by the Cayley-Hamilton theorem $f(\underline{G}) = 0$.

Lemma 2.6: Let $\underline{\epsilon}_t$ be independent identically distributed q×1 random vector with zero expectation and let $\underline{\ell}_i \in \mathbb{R}^q$ be (q+1) constant vectors, i=0,1,...q. Then if $\forall_t$

$$z_t = \sum_{i=1}^{q} \underline{\ell}_i \underline{\epsilon}_{t-i} \qquad (2.4.4)$$

and $\text{cov}(z_t, z_{t-q}) \neq 0$, then z_t is a Moving Average Process of order q and can be written

$$z_t = e_t + \sum_{i=1}^{q} b_i e_{t-i}, \quad \text{where}$$

the e_t's are i.i.d. random variables with zero expectation, and b_i's, i=1,2...p are constants.

A proof of Lemma 2.6 is found in Harrison (1967).

The next theorem shows the representation of a CDLM in the ARMA form.

Theorem 2.7: Given an observable constant DLM

$$Y_t = \underline{F}\,\underline{\theta}_t + \nu_t \quad ; \quad \nu_t \sim [0, V]$$

$$\underline{\theta}_t = \underline{G}\,\underline{\theta}_{t-1} + \underline{w}_t \quad ; \quad \underline{w}_t \sim [0, \underline{W}] \quad ,$$

with V>0, dim $\underline{\theta}$ = p and $\underline{G}$ with eigenvalues $\lambda_1 \ldots \lambda_p$ not necessarily distinct nor necessarily non zero

$$\prod_{i=1}^{p} (1-\lambda_i B)\, Y_t = e_t + \sum_{j=1}^{p} b_j e_{t-j} \text{ , where}$$

the e_t's are i.i.d. random variables with zero mean.

Comment: From sections 2.4.2 and 2.4.3 follows that the class of DLM contains the ARMA class. Since some of the λ's can be zero and so can the b_j's, the class of all ARIMA models is a subset of those described by constant DLM's.

Chapter 3: Normal Dynamic Non-Linear Model.

3.1 Introduction

The D.L.M.'s described in the previous chapter provide a comprehensive framework for the modelling and analysis of a wide class of observed time series.

It is worth pointing out the great flexibility of these models as components of structural forecasting systems and their potential as aids to interpreting the observed behaviour of the process.

Although the Dynamic Bayesian Linear Models of HS(1976) have provided a simple and elegant model formulation and extended the classical models by offering a reformulisation which includes many facilities which are not elegantly available in the classical formulation, they have still some limitations.

Indeed, the N.D.L.M. provides a useful approximation to a more general structure, perhaps after transforming the original data. The problem with this approach are twofold. First we lose the interpretability of the parameters, and secondly when transformations are used the requirements of linearity, constant variance and normality may conflict. In such cases it is desirable that the appropriate non-normal sampling distribution should be used, and transformations avoided.

In this chapter we present a class of Normal Dynamic Non-linear Models which provides a much wider class of applications of the dynamic models to real life problems. The extension to the general case, i.e. non-normal sampling distributions, will be developed in chapter 5.

In section 3.2 we introduce the Normal Weighted Bayesian Model [Ameen-Harrison (1983)], which overcomes some practical disadvantages associated with Dynamic Linear Models. A discount vector replaces the DLM system variance matrix; introduces conceptual simplicity; removes ambiguity and provides a system transition model which is invariant to the measurement scale of independent and control variables.

As a basis for the development of Normal Non-linear Models in section 3.4, a reformulation of the Normal Discount Bayesian Models is discussed in section 3.3.

Finally, section 3.5 is concerned with the applications of Normal Non-linear Models. The construction of models is briefly discussed and the discount concepts used. Some examples of intervention are presented and practical procedures for estimating the observation variance on-line are discussed; in particular, slowly changing variances and variance power laws are presented.

3.2 The Normal Discount Bayesian Models (NDBM).

The Normal Dynamic Linear Models presented in chapter 2 require modellers to be familiar with parametric modelling, state space stochastic difference equations and the Normal probability representation of innovations. The specification of the associated system variance matrices has proved a major obstacle in the applications. In general, people have little quantitative insight for the elements of these matrices. The aim of this section is to introduce a class of Normal Dynamic Bayesian Models which replaces the system variance matrix by a set of discount factors.

3.2.1 Normal Weighted Bayesian Model.

A Normal Weighted Bayesian Model is defined by a quadruple $\{\underline{F}, \underline{G}, V, \underline{H}\}_t$ for each integer $t>0$. Let D_{t-1} be the available data at time t-1 and suppose that the posterior distribution for $\underline{\theta}_{t-1}$ is

$$(\underline{\theta}_{t-1} | D_{t-1}) \sim N[\underline{m}_{t-1}, \underline{C}_{t-1}].$$

Then the model is described by:

(i) observational distribution $(Y_t|\underline{\theta}_t) \sim N[\underline{F}_t\ \underline{\theta}_t,\ V_t]$

(ii) prior distribution: $(\underline{\theta}_t|D_{t-1}) \sim N[\underline{G}_t\ \underline{m}_{t-1},\ \underline{R}_t]$

where:

$$\underline{R}_t = \underline{H}_t\ \underline{C}_{t-1}\ \underline{H}'_t$$

It follows that the predictive distribution is:

$$(Y_t|D_{t-1}) \sim N[\hat{y}_t,\ \hat{Y}_t]$$

where

$$\hat{y}_t = \underline{F}_t\ \underline{G}_t\ \underline{m}_{t-1}$$

$$\hat{Y}_t = \underline{F}_t\ \underline{R}_t\ \underline{F}'_t + V_t$$

Further defining $e_t = y_t - \hat{y}_t$ and $\underline{A}_t = \underline{R}\ \underline{F}'_t/\hat{Y}_t$, the posterior distribution for $\underline{\theta}_t$ is:

$$(\underline{\theta}_t|D_t) \sim N[\underline{m}_t,\ \underline{C}_t]$$

where $\underline{m}_t$ and $\underline{C}_t$ may be calculated recursively by

$$\underline{m}_t = \underline{G}_t\ \underline{m}_{t-1} + \underline{A}_t\ e_t$$

$$\underline{C}_t = [\underline{I} - \underline{A}_t\ \underline{F}_t]\ \underline{R}_t$$

It is worth commenting that the class of Normal D.L.M. is contained in the class of Normal W.B.M. To see this define

$$\underline{W}_t = \underline{H}_t\ \underline{C}_{t-1}\ \underline{H}'_t - \underline{G}_t\ \underline{C}_{t-1}\ \underline{G}'_t.$$

If $\underline{W}_t$ is positive definite for all $t>0$ then the N.D.L.M. $\{\underline{F}, \underline{G}, V, \underline{W}\}_t$ is defined. (see Ameen and Harrison (1983)).

3.2.2 Normal Discount Bayesian Models.

A N.D.B.M. is defined by a quadruple $\{\underline{F}, \underline{G}, V, \underline{\beta}\}_t$ and is a particular N.W.B.M. where always $\underline{H}_t = \underline{B}^{-\frac{1}{2}} \underline{G}_t = \underline{G}_t' \underline{B}^{-\frac{1}{2}}$. Interest will be centered on models structured such that

$$\underline{G}_t = \text{diag}\{\underline{G}_1 \ldots \underline{G}_p\}_t \text{ has a block diagonal form.}$$

Defining $\underline{B}_t = \text{diag}\{\beta_1 I_1 \ldots \beta_p I_p\}$, where:

I_i - is an identity matrix of dimension p_i;

β_i - are the discount factors satisfying $0<\beta_i\leq 1 \ \forall i=1,\ldots p$, and

$\underline{H}_t = \underline{B}^{-\frac{1}{2}} \underline{G}_t' = \underline{G}_t \underline{B}^{-\frac{1}{2}}$

the NDBM is obtained.

We have used these ideas in our applications (section 3.5) and a full account of these models is presented in Ameen and Harrison (1983).

3.2.3 Modified Normal Discount Bayesian Models.

A further class of NWBM's called Modified NDBM's is introduced in this section. The reason for introducing this model class is that there may be a requirement that the system evolution for the different blocks $\underline{G}_i$ be considered as independent of one another. The information from the observation series Y_t is distributed amongst the components of $\underline{\theta}$ introducing parameter covariances. However this does not signify that covariances should be developed over time by discounting or by any other means.

In order to define the modified NDBM consider $(\underline{\theta}_{t-1}/D_{t-1}) \sim N[\underline{m}_{t-1}, \underline{C}_{t-1}]$ where, corresponding to any diagonal block structure for $\underline{G} = \text{diag}\{\underline{G}_1 \ldots \underline{G}_p\}$, at time t-1, the partitioned structure of $\underline{C}_{t-1}$ is $\{\underline{C}_{ij}\}$, of $\underline{R}_t$ is $\{\underline{R}_{ij}\}$. for $i,j=1,2,\ldots p$, and $\underline{B}_t = \text{diag}\{\underline{b}_1, \underline{b}_2 \ldots \underline{b}_p\}$.

Then a modified NDBM $\{\underline{F}, \underline{G}, V, \underline{B}\}_t$ defines $(\underline{\theta}_t|D_{t-1}) \sim N(\underline{G}\,\underline{m}_{t-1}, \underline{R}_t)$ where

$$\underline{R}_{ii} = \underline{b}_i^{-\frac{1}{2}} \underline{G}_i \underline{C}_{ii} \underline{G}_i' \underline{b}_i^{-\frac{1}{2}} \quad i=1,2,\ldots,p,$$

but

$$\underline{R}_{ij} = \underline{G}_i \underline{C}_{ij} \underline{G}_i' \quad \text{for } i\neq j.$$

3.3 Reformulation of the DLM.

In order to obtain an extension of the Normal DLM for the non-linear case we commence by a reformulation of the standard normal analysis.

Suppose we write the NDBM as

(i) Observation distribution $(Y_t|\psi_t) \sim N(\psi_t, V_t)$

(ii) Linear guide relationship: $\psi_t = \underline{F}_t\, \underline{\theta}_t$, where $\underline{\theta}_t$ is a vector of the underlying parameters.

(iii) Prior distribution for ψ_t given the information up to and including time t-1

$$(\psi_t|D_{t-1}) \sim N(\hat{\psi}_t, r)$$

(iv) Finally, the prior distribution for $\underline{\theta}_t$ given D_{t-1}

$$(\underline{\theta}_t|D_{t-1}) \sim N(\hat{\underline{\theta}}_t, \underline{R}_t) \tag{3.3.1}$$

If, in addition, we assume that the likelihood of Y_t given $\underline{\theta}_t$ and ψ_t depends on $\underline{\theta}_t$ only through ψ_t, then using Bayes theorem we obtain

$$\begin{aligned} p(\underline{\theta}_t,\psi_t|D_t) &\propto p(Y_t|\psi_t)\, p(\underline{\theta}_t,\psi_t|D_{t-1}) \\ &\propto p(Y_t|\psi_t)\ p(\psi_t|D_{t-1})\ p(\underline{\theta}_t|\psi_t D_{t-1}) \end{aligned} \tag{3.3.2}$$

Note that the product of the first two terms in the right hand side of (3.3.2) is proportional to the posterior density of ψ_t, so

$$p(\underline{\theta}_t,\psi_t|D_t) \propto p(\psi_t|D_t)\ p(\underline{\theta}_t|\psi_t,D_{t-1}) \tag{3.3.3}$$

The posterior density of ψ_t has the standard form

$$p(\psi_t|D_t) \sim N(m,\sigma_{11})$$

$$\begin{cases} m = \hat{\psi}_t + \dfrac{r}{r+V_t}\,(y_t-\hat{\psi}_t) \\ \sigma_{11} = \dfrac{r}{r+V_t}\ V_t \end{cases}$$

On the other hand the second term in the RHS of (3.3.3) is calculated from the joint prior distribution for $(\psi_t, \underline{\theta}_t)$:

$$\left[\begin{pmatrix} \psi_t \\ \underline{\theta}_t \end{pmatrix} \Bigg| D_{t-1}\right] \sim N\left[\begin{pmatrix} \hat{\psi}_t \\ \hat{\underline{\theta}}_t \end{pmatrix}; \begin{bmatrix} r & \underline{r} \\ \underline{r}' & \underline{R}_t \end{bmatrix}\right]$$

where:

$$\begin{cases} \underline{r} = \mathrm{cov}(\underline{F}_t\ \underline{\theta}_t;\ \underline{\theta}_t / D_{t-1}) = \underline{F}_t\ \underline{R}_t \\ r = \mathrm{var}(\underline{F}_t\ \underline{\theta}_t | D_{t-1}) = \underline{F}_t\ \underline{R}_t\ \underline{F}_t' \end{cases}$$

Using the standard Normal theory it follows that the posterior for $\underline{\theta}_t$ given $\{D_{t-1}, \psi_t\}$ is

$$(\underline{\theta}_t | D_{t-1}, \psi_t) \sim N(\overline{\underline{\theta}}_t, \underline{C}_t)$$

where:

$$\begin{cases} \overline{\underline{\theta}}_t = \hat{\underline{\theta}}_t + \dfrac{\underline{r}'}{r} [\psi_t - \hat{\psi}_t] \\ \underline{C}_t = \underline{R}_t - \underline{r}'\ \underline{r}\ / r \end{cases} \qquad (3.3.4)$$

Now following West, Harrison and Migon (1983), we show that these updating relations coincide with the Kalman filter equations in the Normal linear case. First we note that:

$$\text{(i)} \quad E[\underline{\theta}_t | D_t] = E\left[E[\underline{\theta}_t | D_{t-1}, \psi_t] / D_t\right]$$

$$= E[\overline{\underline{\theta}}_t | D_t]$$

$$= \hat{\underline{\theta}}_t + \frac{\underline{r}'}{r + V_t} [y_t - \hat{y}_t] \qquad (3.3.5)$$

where:

$$\hat{y}_t = E[Y_t | D_{t-1}]$$

$$= E[E(Y_t | \psi_t) / D_{t-1}]$$

$$= E[\psi_t | D_{t-1}] = \hat{\psi}_t$$

$$\text{(ii)}\quad \mathrm{var}(\underline{\theta}_t|D_t) = v\left[E[\underline{\theta}_t|\psi_t,D_{t-1}]/D_t\right] + E\left[v[\underline{\theta}_t/\psi_t D_{t-1}]/D_t\right]$$

$$= \left[\frac{\underline{r}'\underline{r}}{r^2}\sigma_{11}\right] + \left[\underline{R}_t - \frac{\underline{r}'\underline{r}}{r}\right]$$

$$= \underline{R}_t - \underline{r}'\,\underline{r}/(r+V_t) \qquad (3.3.6)$$

Which are the Kalman filter recurrence relations with $A_t = \frac{\underline{r}}{r+V_t}$.

The main conclusions from this sort of analysis, which will help us in the extension to the non-linear normal case are:

(i) the updating of the distribution of the underlying parameter vector $\underline{\theta}_t$ follows from the use of Bayes theorem on ψ_t and from the filtering back of information via the conditional distribution of $\underline{\theta}_t$ given $\underline{\theta}_{t-1}$ and ψ_t;

(ii) from equation (3.3.4) it is clear that $\overline{\underline{\theta}}_t$ depends on D_{t-1} only through the change in the mean of ψ_t, and similarly $\underline{C}_t$ depends only on D_{t-1} through changes in the variance of ψ_t;

(iii) in the normal case the first and second moments completely characterise the probability distribution;

(iv) finally the forecast distribution for $(Y_t|D_{t-1})$ depends on the prior for $\underline{\theta}_t$ only through the parameters of $(\psi_t|D_{t-1})$.

In the coming section we present non-linear analogues for this model where the prior and posterior for the underlying parameters are only partially specified in order to obtain a tractable sequential analysis.

3.4 Normal Dynamic Non-linear Model.

Let $\underline{\theta}$ be a parameter vector with $(\underline{\theta}_{t-1}|D_{t-1}) \sim N[m_{t-1},\underline{C}_{t-1}]$, and consider a specified dynamic model, which may be non-linear, applied to this posterior distribution. The prior distribution for $\underline{\theta}_t$ will be

$$(\underline{\theta}_t|D_{t-1}) \sim N[\hat{\underline{\theta}}_t, \underline{R}_t]$$

Let ψ be a parameter related to $\underline{\theta}$ through a stated "guide relationship".

$$\psi_t = g(\underline{\theta}_t), \quad \text{g-non-linear,}$$

which may be deterministic or stochastic.

The joint prior distribution for $(\psi_t, \underline{\theta}_t)'$ is then defined such that:

$$\begin{bmatrix} \psi_t \\ \underline{\theta}_t \end{bmatrix} \Big| D_{t-1} \sim N \left[\begin{bmatrix} \hat{\psi}_t \\ \hat{\underline{\theta}}_t \end{bmatrix} ; \begin{bmatrix} r & \underline{r} \\ \underline{r}' & \underline{R}_t \end{bmatrix} \right] \tag{3.4.1}$$

where:

$$\hat{\psi}_t = E[g(\underline{\theta}_t)/D_{t-1}]$$

$$r = V[g(\underline{\theta}_t)/D_{t-1}]$$

$$\underline{r} = \mathrm{cov}[g(\underline{\theta}_t),\ \underline{\theta}_t/D_{t-1}] \tag{3.4.2}$$

Comment: In general the marginal distribution of ψ_t does not need to be normal as we see in chapter 5, even though we consider only the normal case in this chapter.

In many non-linear models the higher moments of $\underline{\theta}$ are needed and often we take the marginal distribution of $\underline{\theta}_t$ as normal. In this way we can use the results about the general moments of products of normal random variables (see appendix 3.1) in order to use a Taylor expansion including moments of higher order.

Let the posterior mean and variance of $(\psi_t|D_t)$ be:

$$E(\psi_t|D_t) = m$$

$$V(\psi_t|D_t) = \sigma_{11}$$

and this information is then conveyed back to $\underline{\theta}_t$ as in section 3.3.

The resulting updating equation derived from (3.4.1) are:

$$(\underline{\theta}_t | D_t) \sim N[\underline{m}_t, \underline{C}_t]$$

where: $$\underline{m}_t = \hat{\theta}_t + \underline{r}'(m-\hat{\psi}_t)/r$$

$$\underline{C}_t = \underline{R}_t - \underline{r}'\,\underline{r}(r-\sigma_{11})/r \qquad 3.4.3$$

3.4.2: Model summary and forecasting

For reference purposes the full system of recursions is summarized here:

(i) observational model

$$p(Y_t | \psi_t) \sim N(\psi_t, V_t)$$

(ii) The guide relationship

$$\psi_t = g(\underline{\theta}_t) + \delta, \delta \sim N[0, \sigma_\delta^{\,2}]$$

(iii) Prior distribution

$$(\psi_t | D_{t-1}) \sim N[\hat{\psi}_t; r]$$

$$(\underline{\theta}_t | D_{t-1}) \sim N[\hat{\underline{\theta}}_t, \underline{R}_t]$$

(iv) Posterior distribution

$$(\psi_t | D_t) \sim N[m, \sigma_{11}]$$

$$\begin{cases} m = \hat{\psi}_t + \dfrac{r}{r+V_t}\,(y_t - \hat{\psi}_t) \\ \sigma_{11} = \dfrac{r}{r+V_t}\,V_t \end{cases}$$

(v) Underlying parameter updating

$$(\underset{\sim}{\theta}_t | D_t) \sim N[\underline{m}_t, \underline{C}_t]$$

$$\underline{m}_t = \hat{\underline{\theta}}_t + \underline{r}'(m-\hat{\psi}_t)/r \text{ or } \underline{m}_t = \hat{\underline{\theta}}_t + \frac{\underline{r}'}{r+V_t}\,[y_t - \hat{\psi}_t]$$

$$\underline{C}_t = \underline{R}_t - \underline{r}'\,\underline{r}(r-\sigma_{11})/r^2$$

(vi) Underlying parameter evolution

$$\hat{\underline{\theta}}_t = \underline{G}\ \underline{m}_{t-1}$$

$$\underline{R}_t = \underline{B}_t^{-\frac{1}{2}}\ \underline{G}\ \ \underline{C}_{t-1}\ \underline{G}'\ \underline{B}_t^{-\frac{1}{2}}$$

The predictive distribution for ψ_{t+k} and Y_{t+k}, $k>0$, for forecasting purposes, are available in the following form:

(i) From a practical point of view (Ameen and Harrison (1983)) it is most useful to approximate $(\underline{\theta}_{t+k}/D_t)$ by

$$(\underline{\theta}_{t+k}|D_t) \sim N[\underline{m}_t(k);\ \underline{R}_t(k)], \quad \text{where:}$$

$$\underline{m}_t(k) = \underline{G}\ \underline{m}_t(k-1), \quad k>0$$

$$\underline{R}_t(k) = \underline{B}^{-\frac{1}{2}}\ \underline{G}\ \underline{R}_t(k-1)\ \underline{G}'\ \underline{B}^{-\frac{1}{2}}, \text{ with} \tag{3.4.4}$$

$\underline{m}_t(0) = \underline{m}_t$ and $\underline{C}_t(0) = \underline{C}_t$.

(ii) Hence $(\psi_{t+k}|D_t) \sim N[\hat{\psi}_t(k);\ r_t(k)]$; $k>0$, where:

$$\hat{\psi}_t(k) = E[g(\underline{\theta}_{t+k})|D_t];\ r_t(k) = \text{var}[g(\underline{\theta}_{t+k})|D_t], \text{ and}$$

$\psi_t(0) = \hat{\psi}_t$ and $r_t(0) = r$.

(iii) Finally $p(Y_{t+k}|D_t) \sim N[\hat{y}_t(k);\ \hat{Y}_t(k)]$, where:

$$\hat{y}_t(k) = E\left[E[Y_{t+k}|\psi_{t+k}]|D_t\right]$$

$$= E[\psi_{t+k}|D_t]$$

$$= \hat{\psi}_t(k);\ k>0; \quad \text{and}$$

$$\hat{Y}_t(k) = v\left[E[Y_{t+k}|\psi_{t+k}]|D_t\right] + E\left[\text{var}[Y_{t+k}|\psi_{t+k}]|D_t\right]$$

$$= r_t(k) + V_{t+k}; \quad k>0$$

3.4.3. Normal Non-linear Model with unknown variance

A simple modification of the model provides a tractable learning procedure for unknown variance in the normal case. A number of approaches have been adopted for estimating the observation variance as for example Smith (1977), Smith and West (1983), Ameen and Harrison (1983).

It is assumed that the variance $V = \phi^{-1}$ where ϕ is an unknown constant. The probability model may be rewritten as:

(i) $(Y_t|\psi_t,\phi) \sim N[\psi_t, 1/\phi]$

(ii) $(\psi_t|D_{t-1}) \sim N[\hat{\psi}_t, r/\phi]$

$$(\phi|D_{t-1}) \sim G\left[\frac{\alpha_{t-1}}{2} \; ; \; \frac{n_{t-1}}{2}\right]$$

(iii) $(\underline{\theta}_t|D_{t-1}) \sim N[\underline{m}_t, \underline{C}_{t-1}/\phi]$

where the Gamma pdf has the kernel

$$\exp\left[\left(\frac{\alpha_{t-1}}{2} - 1\right)\log\phi - \frac{n_{t-1}}{2}\phi\right]$$

It follows that the recurrence relationships for the updating procedure and the predictive distribution are exactly as defined in 3.4.2 with the setting $\mathbf{V}_t=1$ and:

(i) Joint prior for $(\psi_t,\underline{\theta}_t)'$

$$\begin{pmatrix}\psi_t \\ \underline{\theta}_t\end{pmatrix}\Bigg| D_{t-1}\phi \sim N\left[\begin{pmatrix}\hat{\psi}_t \\ \hat{\underline{\theta}}_t\end{pmatrix} ; \begin{bmatrix} r & \underline{r} \\ \underline{r}' & \underline{R}_t\end{bmatrix}\phi^{-1}\right]$$

(ii) Posterior distribution for ψ_t, ϕ and $\underline{\theta}_t$

(a) $(\psi_t|D_t,\phi) \sim N[m;\sigma_{11}]$

$$\begin{cases} m = \hat{\psi}_t + \dfrac{r}{r+1}(y_t-\hat{\psi}_t) \\ \sigma_{11} = \dfrac{r}{r+1}\phi^{-1}\end{cases}$$

(b) $(\phi|D_t) \sim G\left(\frac{\alpha_t}{2} ; \frac{n_t}{2}\right)$, where

$$\begin{cases} \alpha_t = \alpha_{t-1} + e_t^2/(r+1) \\ n_t = n_{t-1} + 1 \end{cases}$$

Note that the point estimates of ϕ or $1/\phi$ can then be calculated sequentially; for example $\hat{\sigma}_t^2 = \frac{\alpha_t}{n_t}$ is the posterior harmonic mean of the observational variance.

(c) $(\underline{\theta}_t|D_t,\phi) \sim N[\underline{m}_t, \underline{C}_t/\phi]$

$$\begin{cases} m_t = \hat{\underline{\theta}}_t + \underline{r}'[m-\hat{\psi}_t]/r \\ \underline{C}_t = R_t - \underline{r}'\underline{r}(r-\sigma_{11})/r^2 \end{cases}$$

The marginal posterior of $\underline{\theta}_t$ can be obtained as a multivariate t-student using standard Bayesian analysis.

(d) Finally, the marginal predictive distribution is obtained from the joint distribution $(Y_{t+1},\phi|D_t)$ integrating out ϕ, given:

$$p(Y_{t+1}|D_t) \propto \left[\alpha_t + \frac{(y_{t+1}-\hat{y}_{t+1})^2}{r_{t+1}+1}\right]^{-\frac{n_t}{2}} ,$$

where: $r_{t+1} = \text{var}(\psi_{t+1}|D_t,\phi)$

3.4.4 A practical procedure for estimating the observational variance.

Although the full Bayesian treatment given in 3.4.3 provides a tractable learning procedure for unknown observational variance, practitioners would appreciate some simple but effective robust procedure. In particular, slowly changing variances and power laws are two important aspects in applications. A number of approaches have been adopted for estimating the observational variance, for example Smith and West (1983), Harrison and Stevens (1976), Cantarelis and Johnston (1983) and Leonard and Harrison (1977).

For a NORMAL DLM $(\underline{F},\underline{G},\underline{W},V)_t$, as we have shown in Chapter 2, the predictive distribution is $(Y_t|D_{t-1}) \sim N(\hat{y}_t,\hat{Y}_t)$, where:

$$\hat{Y}_t = \underline{F}_t \, \underline{R} \, \underline{F}_t' + V_t$$

Remembering the definition of $\underline{A}_t$, i.e. $\underline{A}_t = \underline{R} \, \underline{F}'/\hat{Y}_t$, it follows that:

$$V_t = [1 - \underline{F}_t\underline{A}_t]\hat{Y}_t \qquad 3.4.5$$

Using (3.4.5) a simple and effective procedure for variance learning can be obtained.

Defining $d_t^2 = (1 - \underline{F}_t\underline{A}_t)\, e_t^2$, $e_t = y_t - \hat{y}_t$, then $E(d_t^2/D_{t-1}, V) = V$. For an unknown constant V's natural estimate is

$$\hat{V}_t = \frac{x_t}{n_t}$$

where:

$$\begin{cases} x_t = x_{t-1} + d_t^2 \\ n_t = n_{t-1} + 1 \end{cases} \qquad 3.4.6$$

In order to robustify the procedure we suggest the practical method:

$d_t^2 = (1 - \underline{F}\,\underline{A}_t)\min[e_t^2;\, k\,\hat{Y}_t]$ where k is generally chosen in the interval [4,6].

Initial values for (x_0,n_0) in (3.4.6) may be chosen such that $\hat{V}_0 = \frac{x_0}{n_0}$ is a point estimate for V and n_0 is the accuracy expressed in terms of a number of degrees of freedom or of equivalent observations.

Cases in which it is suspected that V varies slowly over time, a discount factor can be introduced so that

$$\begin{cases} x_t = \beta_v \, x_{t-1} + d_t^2 \\ n_t = \beta_v \, n_{t-1} + 1 \end{cases} \qquad 3.4.7$$

Since ϕ_t has a skew distribution it is wise to choose $.95<\beta<1$. Further, if the initial prior of the parameters vector $\underline{\theta}$ is vague, then it is recommended that variance learning commences at time $(p+1)$ where p is the vector dimension.

Stevens (1974) gives a theoretical basis for the popular empirical variance law which is often applied in stock control. With positive observations, the law is:

$$V_t = a\, \hat{y}_t^{\,2b}$$

Empirically a value of $b = .75$ is often appropriate. An estimate of a, $\hat{a}_t$, is then derived from

$$z_t = \beta_v z_{t-1} + d_t^2/\hat{y}_t^{\,1\cdot 5}$$

$$n_t = \beta_v n_{t-1} + 1$$

$$\hat{a}_t = \frac{z_t}{n_t} \qquad 3.4.8$$

Future estimates of V are $\hat{V}_{t+k} = \hat{a}_{t+k}\{E[Y_{t+k}|D_t]\}^{1\cdot 5}$ If the power is unknown and the variance law written $V_t = a\,\hat{y}_t^{\,2b}$ the support for (a,b) may be defined

$$S(a,b|D_t) = S(a,b|D_{t-1}) + \left[\frac{d_t^2}{a\hat{y}_t^{\,2b}} - \log a - 2\,b\,\log\hat{y}_t\right]/2$$

The best supported values may be used as in Harrison and Pearce (1971). Ameen and Harrison (1983) recommend that with such a method: $d_t^2 = (1-\underline{F}_t\underline{A}_t)\min(e_t^2, 4\,\hat{Y}_t)$, $\beta \geq .98$, and that strong prior support must be applied.

It is worth pointing out that in our case $\underline{F}_t$ is equal to the scalar 1 and $\underline{A}_t$ is just $\frac{r}{r+V_t}$.

3.5. The Seasonal Growth Multiplicative Model.

3.5.1 General.

In previous studies, Harrison (1965) and Harrison-Scott (1965) have pointed out the importance of the multiplicative model formulation for analysing seasonal data in forms of separate trend and seasonal components.

Our dynamic non-linear model, when applied to the seasonal growth multiplicative case, avoids the inconvenience of transforming the data. The practitioners may operate with parameters relating to the current level, growth rate and seasonal effects rather than to transformed mixtures of these parameters. For example a linear growth on original scale cannot be easily obtained when a logarithmic transformation is taken - the linear growth component then produces an exponential growth in the original observations.

Another point we would like to focus on is the often recommended transformation in order to apply stationary time series theory. The work with $X_t = Y_t^{1-p}$, the transformed series is a clear case of mathematical convenience leading the modeller to adopt an undesirable model whose meaning is not at all clear to the decision maker. With the Bayesian approach there is no need to operate with transformed variables nor to lose interpretability.

Furthermore, it seems to us a quite serious drawback to transformed variables the fact that they may not be robust to departures in the assumptions on which they are derived. For example if a logarithmic transformation is taken and a zero or near zero observation occur then the value of $x_t = \log y_t$ is $-\infty$. This means that protection rules need to be introduced to guard against such occurrences. In general there may be a danger of assigning the much importance to small values, say at a seasonal trough, compared to the important values which occur at seasonal peaks.

3.5.2 Notation and Definitions

Let the vector of the underlying parameter $\underline{\theta}_t$ be

$$\underline{\theta}_t = \begin{pmatrix} \underline{\pi} \\ \underline{\rho} \end{pmatrix}_t , \quad \text{where:} \quad \underline{\pi}_t = \begin{pmatrix} \mu \\ \eta \end{pmatrix}_t \text{ and } \underline{\rho} = (\rho_1 \dots \rho_r)'$$

The vector $\underline{\pi}_t$ represents the trend component, with μ being the level and η the growth; and $\underline{\rho}$ is the vector of seasonal factors.

The guide relationship is defined as:

$$\psi_t = g(\underline{\theta}) = (\underline{F}\ \underline{\pi}_t) * (1 + \underline{H}\ \underline{\rho}) \qquad 3.5.1$$

where: $\underline{F} = (1,0)$ and $\underline{H}$ $(1,0,1,0,\ldots,1,0)$.

We describe the system evolution by the structure:

(i) systematic evolution - let $\underline{G}$ be a $(T+2) \times (T+2)$ matrix.

$$\underline{G} = \mathrm{diag}\{\underline{G}_1;\ \underline{G}_2,\ldots,\underline{G}_{\left[\frac{T}{2}\right]+1}\}$$

where $\underline{G} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $\underline{G}_{k+1} = \begin{bmatrix} \cos(kw) & \sin(kw) \\ -\sin(kw) & \cos(kw) \end{bmatrix}$

for $k=1,2,\ldots\left[\frac{T}{2}\right]$; $w = \frac{2\pi}{T}$.

(ii) Let $\underline{B}$ be a $(T+2) \times (T+2)$ matrix of discount factors defined as:

$$\underline{B} = \mathrm{diag}\{\underline{B}_1;\ \underline{B}_2,\ldots,\underline{B}_{\left[\frac{T}{2}\right]+1}\}$$

where: $\underline{B}_1 = \beta_1\ I_2$ and $\underline{B}_{k+1} = \beta_k\ I_2$, $k=1,\ldots \frac{T}{2}$ and I_2 - identity matrix of dimension 2, with $0<\beta_i\leq 1$, and often $\beta_{k+1} = \beta_2$ for all $k=1,\ldots\left[\frac{T}{2}\right]$

3.5.3 The multiplicative model.

(i) Observational distribution

$$(Y_t|\psi_t) \sim N(\psi_t,V_t)$$

where $V_t\ \alpha\ a\ \hat{y}_t^{\,b}$, with a learning procedure for a as discussed in section 3.4.4

(ii) Prior distribution for $\underline{\theta}_t$

Given the information up to and including time t-1,

$$(\underline{\theta}_{t-1} | D_{t-1}) \sim N\left[\underline{m}_{t-1}, \underline{C}_{t-1}\right]$$

where: $\underline{\theta}_t = \begin{pmatrix} \underline{\pi} \\ \underline{\rho} \end{pmatrix}_t$; $\underline{m}_{t-1} = \begin{pmatrix} \underline{m}_1 \\ \underline{m}_2 \end{pmatrix}$ and $\underline{C}_{t-1} = \begin{bmatrix} \underline{C}_{11} & \underline{C}_{12} \\ \underline{C}_{21} & \underline{C}_{22} \end{bmatrix}$

are partitioned to represent the trend and seasonal components.

Let $\underline{\Sigma} = \underline{G}\,\underline{C}_{t-1}\,\underline{G}'$ be in a corresponding partitioned form. Then define $\underline{R}_{ii} = \underline{B}_i^{-\frac{1}{2}}\,\underline{\Sigma}_{ii}\,\underline{B}_i^{-\frac{1}{2}}$ and, for $i \neq j$, $\underline{R}_{ij} = \underline{\Sigma}_{ij}$ (the modified NDBM).

(iii) Let $X_t = \underline{F}\,\underline{\pi}_t$ and $U_t = \underline{1} + \underline{H}\,\underline{\rho}_t$, then it follows that:

$$\left(\begin{matrix} X_t \\ U_t \end{matrix} \middle| D_{t-1}\right) \sim N\left\{\begin{pmatrix} \bar{x}_t \\ \bar{u}_t \end{pmatrix}; \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}\right\}$$

where: $\bar{x} = \underline{F}\,\underline{m}_1$

$\bar{u} = \underline{H}\,\underline{m}_2 + 1$ and

$v_{11} = \underline{F}\,\underline{R}_{11}\,\underline{F}'$; $v_{12} = \underline{F}\,\underline{R}_{12}\,\underline{H}' = v_{21}$

$v_{22} = \underline{H}\,\underline{R}_{22}\,\underline{H}'$

The joint prior distribution for $(\psi, \underline{\theta})_t$ is

$$\begin{pmatrix} \psi \\ \underline{\theta} \end{pmatrix}_t \Bigg| D_{t-1} \sim N\left[\begin{pmatrix} \hat{\psi}_t \\ \hat{\underline{\theta}}_t \end{pmatrix}; \begin{bmatrix} r & \underline{r} \\ \underline{r}' & \underline{R}_t \end{bmatrix}\right]$$

where:

$$\begin{cases} \hat{\psi}_t = E[\psi_t | D_{t-1}] = E[X_t\,U_t | D_{t-1}] = \bar{x}_t\,\bar{u}_t + v_{12} \\ r = v[\psi_t | D_{t-1}] = V_t + v_{11}\,v_{22} + v_{12}^2 + \bar{x}^2\,v_{22} + \bar{u}^2\,v_{11} + 2\bar{x}\,\bar{u}\,v_{12} \\ \underline{r} = \mathrm{cov}(\psi_t, \underline{\theta}_t | D_{t-1}) = \bar{x}\,\mathrm{cov}(U, \underline{\theta}_t | D_{t-1}) + \bar{u}\,\mathrm{cov}(X, \underline{\theta}_t | D_{t-1}) \end{cases}$$

(iv) The updating

(a) The posterior distribution of ψ_t is

$$(\psi_t \mid D_{t-1}, y_t) \sim N(m, \sigma_{11})$$

where:

$$\begin{cases} m = \hat{\psi}_t + \dfrac{r}{r+V_t} e_t, \; e_t = y_t - \hat{y}_t \\ \sigma_{11} = \dfrac{r}{r+V_t} V_t \end{cases}$$

and,

(b) The posterior for $\underline{\theta}_t$:

$$(\underline{\theta}_t \mid D_{t-1}, \psi_t) \sim N(\underline{m}_t, \underline{C}_t)$$

where

$$\begin{cases} \underline{m}_t = \underline{\hat{\theta}}_t + r' \; [\underline{m} - \hat{\psi}_t]/r \\ \underline{C}_t = \underline{R}_t - \underline{r}' \; \underline{r} \; [r - \sigma_{11}]/r^2 \end{cases}$$

3.5.4 Examples

(a) Artificial Data

For an assessment of the effectiveness of the model an artificial data series is used comprising 108 months observations. The generating model, which is described in the Appendix 3.2, combines a linear trend dynamic model and a 12 period seasonal effect derived from a single first harmonic.

The model is defined as:

$$\underline{G} = \text{diag}\left\{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}; \begin{bmatrix} \cos(w) & \sin(w) \\ -\sin(w) & \cos(w) \end{bmatrix}\right\}, \text{ with } w = \frac{\pi}{6}$$

and $\beta = \text{diag}\{\beta_1 \, I_2, \beta_2 \, I_2\}$, $\beta_1 = .99$ and $\beta_2 = .91$ and $V_t = 2500 \; \forall t$.
A relatively vague prior specification was made:

$$(\underline{\theta}_0|D_0) \sim N\left[\begin{pmatrix}450.0\\5.0\\0.0\\0.0\end{pmatrix} ; \left[\begin{array}{cc|c} 10000 & 0 & \underline{0} \\ 0 & 4.0 & \\ \hline \multicolumn{2}{c|}{\underline{0}} & .1\, I_2 \end{array}\right]\right] , \text{ which}$$

corresponds to a specification of no seasonal pattern! In figure 3.5.1 we can see the observations and the one-step-ahead forecast.

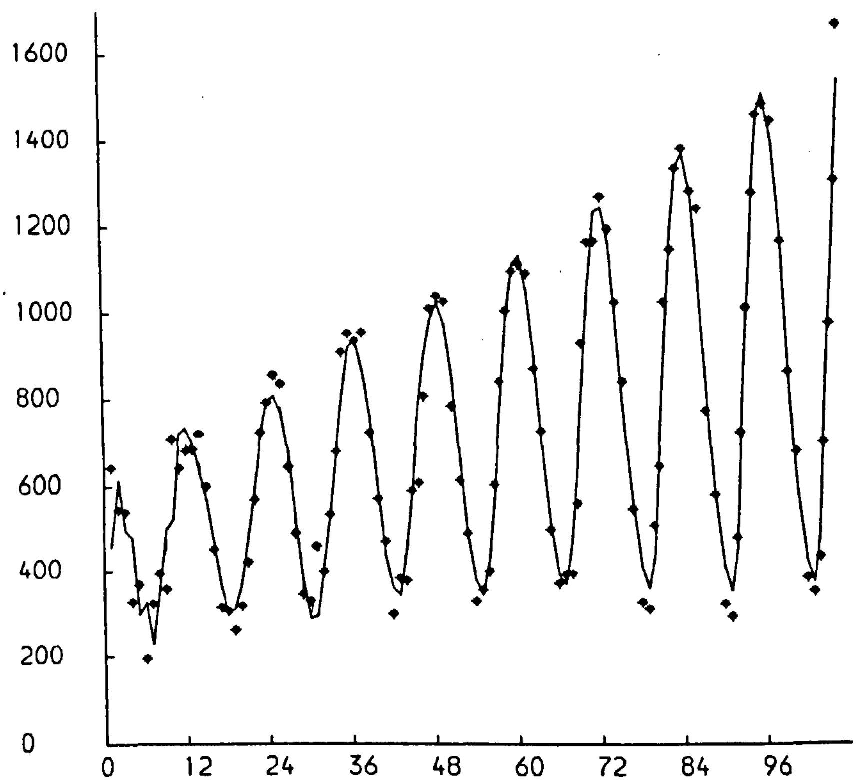


Fig. 3.5.1. Artificial data (*) and one step ahead forecast ($\sim$) $\beta_1 = .99$ $\beta_2 = .91$.

The performance in terms of the mean absolute desviation (MAD) for each of the last six years is:

year	1	2	3	4	5	6
MAD	53.8	33.3	49.7	47.0	40.9	46.0

The observational variance was estimated sequentially given an estimate of 2510.7. It is worth pointing out that no protection against outliers is used in the variance estimation.

The final estimates for level and growth are

	mean	variance
level	988.12	145.55
growth	5.09	.06

(b) The Turkey Poult Data

The main characteristics of this data set are:

(i) There is clearly a growth in sales over the 10 years and an annual seasonal pattern.

(ii) Many events, as discussed in Ameen and Harrison (1981) took place, such as changes in feed price and some promotional campaigns.

The desirability of incorporating informatiom on these events either by modelling them or through intervention emphasizes the importance of a flexible operational system. For illustration purposes an intervention on the level was made at observation 29. The following seasonal linear growth was applied to the original data:

$$\underline{F} = (1,0) \ ; \ \underline{H} = (1,0\ldots1.0), \ \underline{G} = \text{diag}\{G_1\ldots G_6\}$$

$$\underline{\beta} = \text{diag}(\beta_1 I_2; \ \beta_2 \ I_{10}\}$$

$$\underline{G} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \ \underline{G}_{k+1} = \begin{bmatrix} \cos kw & \sin kw \\ -\sin kw & \cos kw \end{bmatrix}, \ \begin{matrix} w = \pi/6 \\ k = 1,\ldots 6. \end{matrix}$$

The initial information was:

	mean	variance
level	80	400.0
growth	1.0	.0625
seas. comp.	0.0	.18

and the pair of discount factor $(\beta_1 \beta_2) = (.95, .93)$.

The power law, with b=1, was estimated on line using a discount factor $\beta_v = .98$

From Fig. 3.5.2 it is evident that, like most practical series, this data is far from stationarity. Variables such feed price might beneficially be introduced but they also need to be estimated. Many of the events which contribute large variation could be anticipated but can only be subjectively described.

The performance of the non-linear Bayesian model in this case was quite good. In terms of the MAD for each year we have:

year	6	7	8	9	10	11
MAD	17.5	18.8	21.2	35.7	40.0	23.9*

*This figure is the mean for nine months

The average run length of errors with same sign is 2.16, the MSE for the last 5 years is 1236.87 and the percentual MAD is 23.5%.

At observation number 29 we drop the level by 30 units and raise the discount factor to some power in order to represent our extra uncertainty about the new level. This means that the model is very adaptive with respect to the trend, and in order to keep the seasonal block unchanged a modified discount model was used.

Finally it is worth pointing out that the discount factors were not chosen to optimise a performance criterion but are rounded figures which were thought to be appropriate.

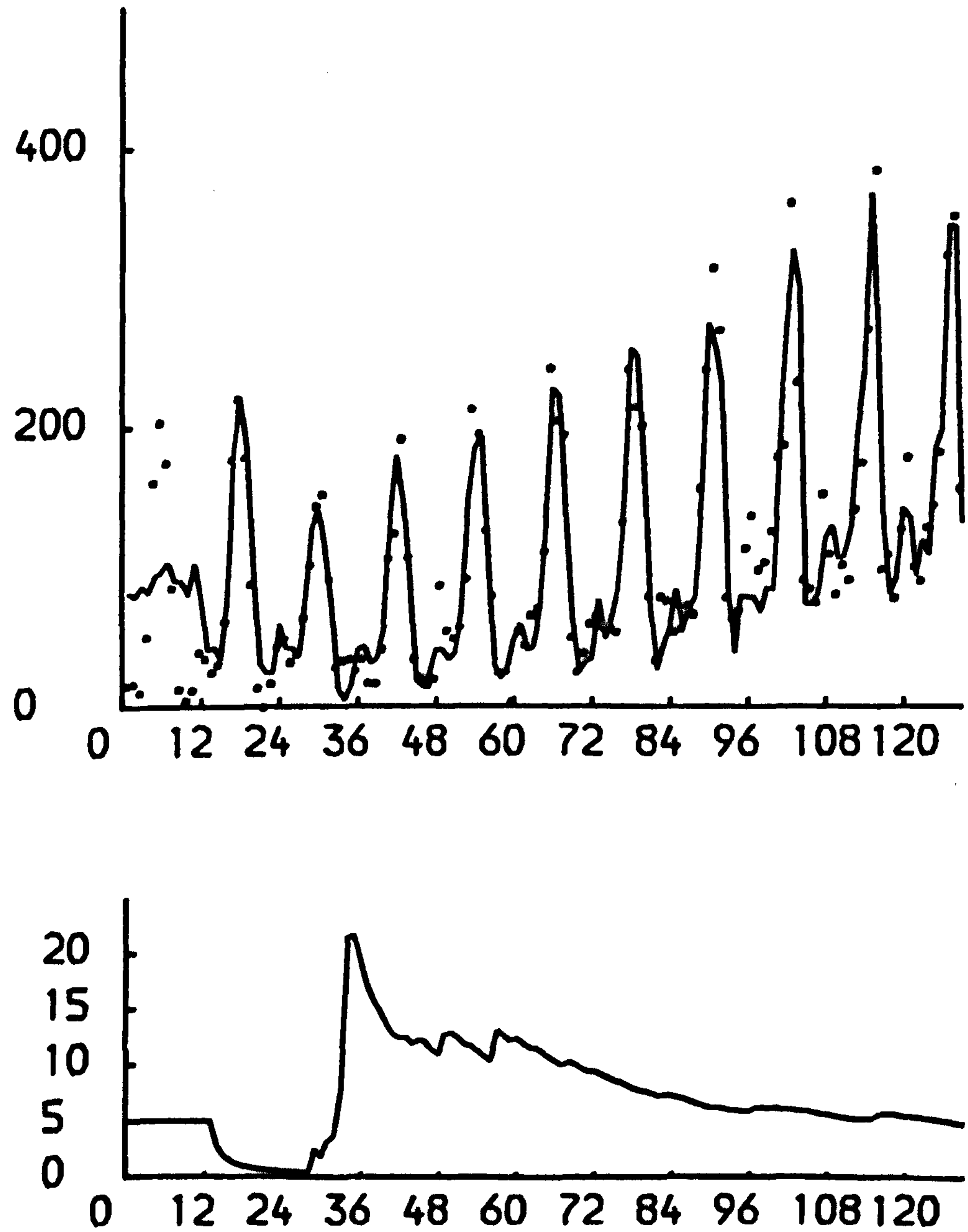


Fig. 3.5.2. Turkey data (*); one-step-ahead forecast and variance learning. $\beta_1 = .95$ and $\beta_2 = .93$.

(c) An example of transfer response: Champagne Data.

The main purpose to present this example is the opportunity to introduce an application of transfer response. This data corresponds to the monthly champaign sales of a French Company, in millions of bottles, from Jan. 1964 up to Sept. 1972.

Since the demand for champagne follows a marked seasonal pattern during the course of the year a linear growth seasonal model in multiplicative form was fitted to the data.

The non-linear model is applied with $\underline{F} = (1,0)$, $\underline{H} = (1,0,\ldots 1,0)$ and $\underline{G} = \text{diag}(\underline{G}_1\ \underline{G}_2 \ldots \underline{G}_7)$, where

$$\underline{G}_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \underline{G}_{k+1} = \begin{bmatrix} \cos(kw) & \sin(kw) \\ -\sin(kw) & \cos(kw) \end{bmatrix}, \quad k=1,\ldots 6 \text{ and } w = \frac{\pi}{6}.$$

A vague initial prior: $(\theta_0|D_0) \sim N[(28, .10,.0\ldots 0)'];\text{diag}(2500.0, .04, .10\ I_{12})$ was adopted. The discount matrix was set as $\underline{B} = \text{diag}(.93\ I_2, .97\ I_{12})$ and the residual variance was estimated on-line.

After the fitting of the above model the residuals for the period 72 to 78 exhibit a quite uncommon structure which might be interpreted as an anticipation of demand. This extra demand at observation 72 (Dec. 69) is followed by a return to the original level as suggested by fig. 3.5.3.

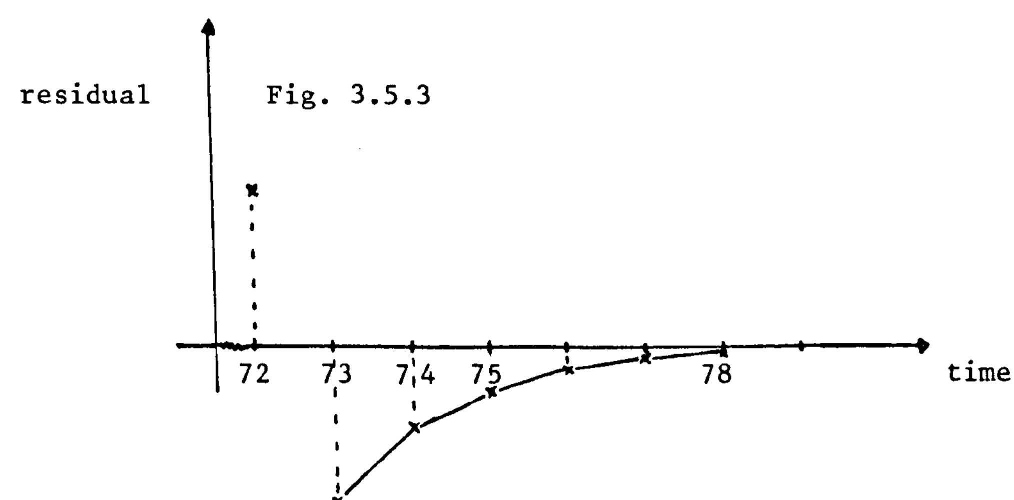


Fig. 3.5.3

Mathematically this effect could be modelled as:

$$E_t = \lambda E_{t-1} + \gamma x_t; \; x_t = \begin{cases} 1 \text{ if } t = t_0 \\ 0 \text{ if } t \neq t_0 \end{cases}; \; \lambda\text{-known}; \; \gamma \sim N[\bar{\gamma}, V_\gamma]$$

and the guide relationship (5.1) is rewritten as: $\psi_t^* = \begin{cases} \psi_t + E_t \text{ if } t \le t_0 \\ \psi_t - E_t \text{ if } t > t_0 \end{cases}$

This transfer response model was added to the basic model in order to improve it. Fig. 3.5.4 shows the one month ahead mean forecast and the actual figures.

In this application of transfer response λ was assumed known as .90 and $\gamma \sim N[20,400]$. The MSE for the last 6 years was 41.65 and percentual MAD 11.2%.

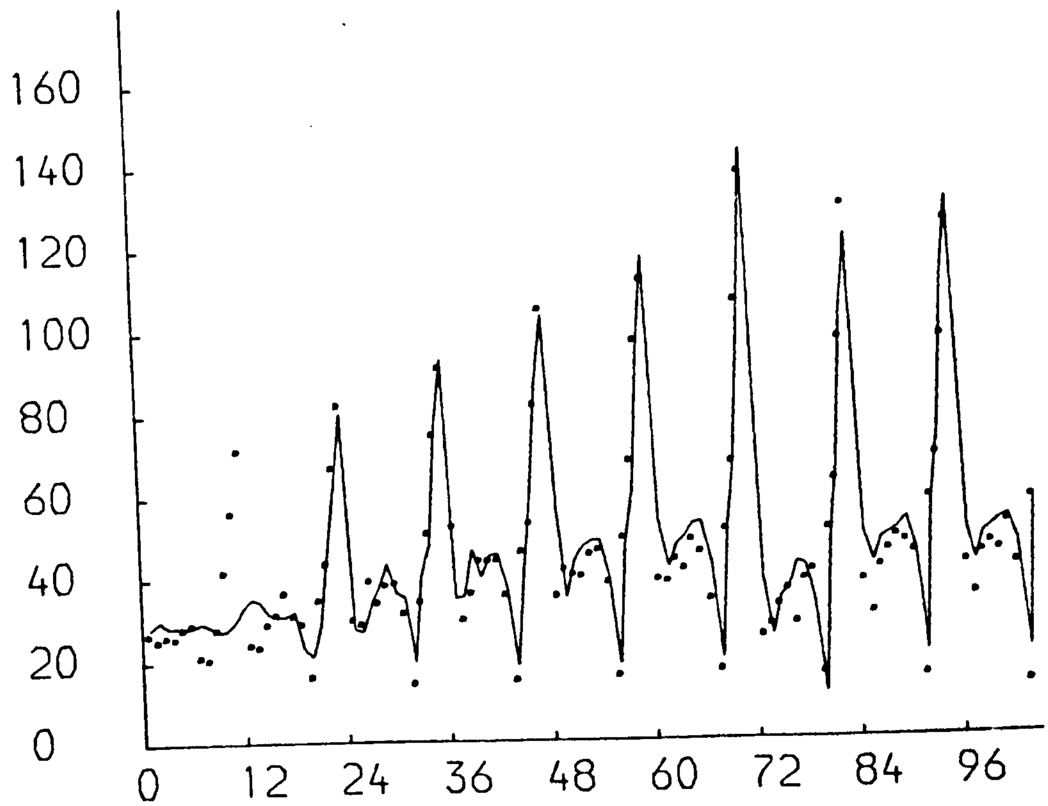


Fig. 3.5.4. Champagne data (*) and one-step-ahead forecast $\beta_1 = .93$; $\beta_2 = .97$

As an assessment of the effect of the transfer response we have estimated the model with and without this component and the MAD were as follows:

Year	3	4	5	6	7	8
Without intervention	4.2	4.8	4.9	10.0	3.8	2.4
With intervention	4.2	4.8	3.8	5.3	4.5	2.6

These figures were obtained with a fixed variance (to avoid interaction from this latter effect), and the final results presented in fig. 3.5.4 include variance learning.

Appendix 3.1 General Moments for the Multivariate Normal Distribution.

(i) General Moments about the mean.

Let $\underline{X} = (X_1 \ldots X_n)' \sim N(\underline{0}, \underline{C})$ so that without loss of generality $E(\underline{X}) = 0$.

The moment generating function of $\underline{X}$ is:

$$\phi(\underline{t}) = \exp\{-\underline{t}\ \underline{C}'\ \underline{t}/2\}$$

with the expansion

$$\phi(\underline{t}) = \sum_{v=0}^{\infty} (-\underline{t}'\ \underline{C}\ \underline{t})^v / v!$$

Theorem: All odd order moments about the mean are zero, and the even order moments about the mean are:

$$\mu_k\{1,2,\ldots 2k\} = E(X_1 X_2 \ldots X_{2k})$$

$$= \Sigma(c_{ij}\ c_{kl} \ldots c_{x2}) \text{ where the}$$

sum is taken over all permutations of the subscripts giving $\frac{2k!}{2^k\ k!}$ terms in the sum, each being the product of k covariances.

Proof: We know it is true for k=1. By finite induction

$$\mu_{2k} = E(X_1 \ldots X_{2k}) = \frac{1}{2^k\ k!} \frac{\partial^{2k}}{\partial t_1 \ldots \partial t_{2k}} (\underline{t}'\ \underline{C}\ \underline{t})^k \Big|_{\underline{t}=0}$$

Since $\underline{t}'\ \underline{C}\ \underline{t} = \sum_i \sum_j c_{ij}\ t_i\ t_j$ and

$$\frac{1}{2kk!} \frac{\partial(\underline{t}'\ \underline{C}\ \underline{t})^k}{\partial t_i} = \frac{2k}{k!} (\underline{t}'\ \underline{C}\ \underline{t})^{k-1} \sum_j C_{ij}\ t_j$$

$$= \frac{(\underline{t}'\ \underline{C}\ \underline{t})^{k-1}}{2^{k-1}(k-1)!} \sum_j C_{ij}\ t_j$$

We have:

$$\mu_{2k} = \sum_j c_{ij}\ \mu_{2(k-1)}(\widetilde{ij}) \quad \text{where}$$

$\mu_{2k-1}(\widetilde{i,j})$ is the $2(k-1)^{th}$ moment not involving one i and one j term.

Hence the result follows by induction.

Example: Fourth order moments about the mean.

$$E(X_i^4) = 3\ c_{ii}^2$$

$$E(X_i^3\ X_j) = 3\ c_{ii}\ c_{ij}$$

$$E\ X_i^2\ X_j^2 = c_{ii}\ c_{jj} + 2\ c_{ij}^2$$

$$E\ X_i^2\ X_j\ X_k = c_{ii}\ c_{jk} + 2\ c_{ij}\ c_{ik}$$

$$E\ X_i\ X_j\ X_k\ X_\ell = c_{ij}\ c_{kl} + c_{ik}\ c_{kl} + c_{il}\ c_{jk}$$

(ii) Ordinary moments of Multivariate Normal.

Let Ω_n be a set of n r.v.'s $\{X_1 \ldots X_n\}$ and $Q_i = \{(\Omega - A_i; A_i)$ such that A_i is any subset of i of the elements$\}$

Define $G_{n,i} = \sum_{Q_i} \mu_{A_i}\ P_{A_i}$ $\quad i=1,\ldots n-1$ with

$$G_{n,n} = \overset{n}{\pi}\ \mu A_i$$

Lemma: $\pi\ X_i = \sum_{i=0}^{n} G_{n,i}$

Definition: $S_{n,i} = E[G_{n,i}] = \sum_{Q_i} \mu_{A_i} c.E[PA_i]$

$E(P_{A_i})$ are given from $\underline{i}$

Theorem: $E\left[\prod_{i=1}^{n} X_i\right] = \sum_{i=0}^{n} S_{n,i}$

Appendix 3.2: Artificial Data Generation.

A sample of 108 observations was generated in the following way:

(i) Let $\begin{pmatrix} \underline{\pi}_t \\ \underline{\rho}_t \end{pmatrix}$ be defined as:

$$\begin{pmatrix} \underline{\pi} \\ \underline{\rho} \end{pmatrix}_t = \underline{G} \begin{pmatrix} \underline{\pi} \\ \underline{\rho} \end{pmatrix}_{t-1} + \underline{w}_t \ ; \text{ where } \underline{G} = \begin{bmatrix} \underline{G}_1 & 0 \\ \underline{0} & \underline{G}_2 \end{bmatrix} \text{ with}$$

$$\underline{G}_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \text{ and } \underline{G}_2 = \begin{bmatrix} \cos(\pi/6) & \sin(\pi/6) \\ -\sin(\pi/6) & \cos(\pi/6) \end{bmatrix} \ ; \text{ and}$$

$$\underline{w}_t \sim IN[0, \underline{W}]$$

The initial setting was:

$$\underline{\pi}_t = (450,5)'; \ \underline{\rho} = (144, .11)' \text{ and}$$

$$\underline{w}_1 = \text{diag}(13.16;.019), \ \underline{w}_2 = .00005 \ I_2 \text{ and}$$

$$\underline{W} = \text{diag}\{\underline{w}_1, \underline{w}_2\}$$

$$\underline{V} = 2500.0$$

(ii) Let X and U be defined as: $X = \underline{F}\ \theta$ and $U = \underline{H}\ \underline{\rho} + 1.0$, $\underline{F} = (1,0) = \underline{H}$, so that

$$\left[\begin{pmatrix} X \\ U \end{pmatrix}_t \Big/ \begin{pmatrix} \theta \\ \rho \end{pmatrix}_{t-1}\right] \text{ is } N\left[\begin{pmatrix} \bar{x} \\ \bar{u} \end{pmatrix} ; \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}\right]$$

(iii) Obtain the joint distribution

$$\left[\begin{pmatrix} \underline{\theta}_t \\ \underline{\rho}_t \\ Y_t \end{pmatrix} \Big/ \begin{pmatrix} \theta \\ \hline \rho \end{pmatrix}_{t-1}\right] \sim N\left[\begin{pmatrix} G\begin{pmatrix} \pi \\ \hline \rho \end{pmatrix}_{t-1} \\ \bar{x}\ \bar{u} + v_{12} \end{pmatrix} ; \left[\begin{array}{cc|c} \multicolumn{2}{c|}{\underline{W}} & \underline{\Sigma}_{13} \\ & & \underline{\Sigma}_{23} \\ \hline \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{array}\right]\right] ;$$

(iv) Finally the conditional distribution of $\left(Y_t / \begin{pmatrix} \pi \\ \rho \end{pmatrix}_t\right)$ is obtained as:

$$\left[Y_t / \begin{pmatrix} \pi \\ \hline \rho \end{pmatrix}_{t-1}\right] \sim N\left[v_{12} + \bar{x}\ \bar{u} + A_t \begin{pmatrix} \pi \\ \hline \rho \end{pmatrix}_t - G\begin{pmatrix} \pi \\ \rho \end{pmatrix}_{t-1}, \Sigma_{33} - A_t\ \underline{w}\ A_t'\right]$$

(v) Using the conditional data in (iv) a sample of 108 observations were generated.

Chapter 4: An Application of Dynamic Non-Linear Bayesian Models to T.V. Advertising.

4.1: Introduction.

This chapter comprises a report on some work developed for a Market Research Company [Millward-Brown Co.]. The problem is one of exploring the relationship between advertising awareness and advertising weight. The objective is to evaluate the effectiveness of TV-Advertising and to provide information on the response and decay of advertising effects.

The descriptive model relating TV-Advertising and consumer awareness has been successfully used for two years both for retrospective and prospective analysis.

Previous analyses of advertising data have focused on the relation between sales and advertising, and discussed the existence of feedback causal relationship. The classical method of transfer functions has been used as in Bhattacharrya et al (1982) and Caines et al (1977). A good survey of the econometric models applied to the measurement of the duration of advertising effect was done by Clarke, D.G. (1976). Mahajan, V. et al (1980) have developed models which can adjust automatically to changing data patterns in the postulated response function. In particular, Broadbent (1979) has applied lagged polynomial econometric models in fitting the relationship between awareness and advertising.

Our approach to this application involves the concept of stochastic transfer response in the context of dynamic linear and non-linear models. The performance of these models are assessed and discussed. The method developed in Chapter 3 is used with the evolution of state parameters as a DLM. The performance of these models are assessed and discussed.

In the coming sections we discuss aspects of the data and state the bases of the models. Section 2 shows the objectives of the study and discusses some aspects of the two measurements. The basic assumptions as well as the facilities required by the Millward Brown are discussed in Section 3.

Section 4 shows how extra facilities and improvements can be obtained within the class of Dynamic linear models, Bayesian Forecasting, [Harrison-Stevens (1971-1976)] and Kalman Filtering, [Kalman (1963)]. These models are extentions of Broadbent (1979).

Section 5 states some conclusions derived from the basic assumptions. The effect of advertising and of consumer memory ageing are modelled and combined to produce a basic guide relationship on which the non-linear models are based. The observational model is discussed and measures for assessing the effectiveness of advertising and for comparing different advertisements are presented.

Section 6 shows applications of the models and gives a comparison between the performance of the linear and non-linear models over a set of fast moving consumer goods. The general conclusion is that the non-linear models provide a significantly better description of the relationship and give marked improvements in forecast performance.

4.2: The Advertising Project.

This project was developed for the Millward Brown Company which has various clients in about 16 different markets of fast moving consumer goods such as lager, chocolate bar, breakfast cereal etc.

In any market, where a company has one or more brands supported by a substantial amount of advertising, at least some of which is on television (T.V.), there is considerable benefit to be derived from setting up a vehicle for continuously monitoring the consumer response to advertising, say Millward Brown (1983). Such data is valuable for reviewing previous advertising; for correcting the tendency of many current pre-testing methods to give a systematically misleading feed-back to the creative team; and for leading to more rational advertising decisions. Measures of brand image, intention to buy and so on are often derived from the data. However, this thesis is concerned with the relationship between consumer awareness and television advertising.

It is worth noting that such advertising usually takes place in bursts or waves, so that the population consumer awareness varies considerably, creating a dynamic situation which merits frequent and perhaps continual study. For some products a sample of around one hundred people in selected regions of the country is taken by Millward Brown in many weeks.

4.2.1: Objectives of the Study

The main objectives of such analysis are:

(i) to define a relationship between TV advertising and consumer's awareness;

(ii) to separate the effects of advertising spending from that of advertisements' content, and so to evaluate the effectiveness of the advertising;

(iii) to assess the merits of different advertising campaigns;

and

(iv) to provide information on the response and decay of advertising effects in order to help in scheduling decisions.

4.2.2: The Measurements

The two measures we deal with are the percentage of people aware of a brand and the advertising weight.

Advertising weight:

T.V. advertising is measured by JICTAR as weekly T.V. ratings (T.V.R.'s) over populations which are usually different from those in the consumer survey. They are taken over the same T.V. areas. A common measure for commercials of different lengths is required. This is usually standardised on 30 seconds, increasing the T.V.R.'s for longer advertisements and decreasing them for shorter ones in proportion to their costs. Further discussion of T.V.R.'s may be found in Broadbent (1979).

Consumer Awareness:

Awareness may be measured in a variety of ways. Each person may be shown a list of brands and each respondent asked "which of these have you seen advertised on T.V. recently?". One awareness measure for a brand may be based upon the sampled number and the number of positive responses for that brand.

4.3: System Model - Basic Assumptions.

It is important to emphasize here that the models were developed having in mind some desired facilities in order to attain the objectives in 4.2.1, and are based on stated hypotheses. In this section we explore these aspects.

4.3.1 Facilities required.

(i) a capability for handing weeks where no sample data is collected, i.e. missing observations;

(ii) a means of including subjective information, so that:

a) initially modelling and prediction can begin with no data;

b) at any time the parameter estimates and their associated uncertainties can be changed;

(iii) an ability to respond quickly to signalled changes in advertisement and to measure the effect of such changes;

(iv) a "what/if" facility for obtaining an estimate of a projected advertising campaign and also for assessing the model adequacy;

(v) a filtering facility for estimating the underlying weekly awareness for week t based upon data up to and possibly including week t+5. This filters the random variation and aims to give a smoother retrospective picture of underlying awareness.

4.3.2: The Basic Assumptions.

The models referred above were based on the following propositions:

(i) A diminishing return rate from increased advertising. Consider a moment of time at which the awareness effect E is derived from an advertising stock a, then the hypothesis of diminishing returns expresses the relationship as:

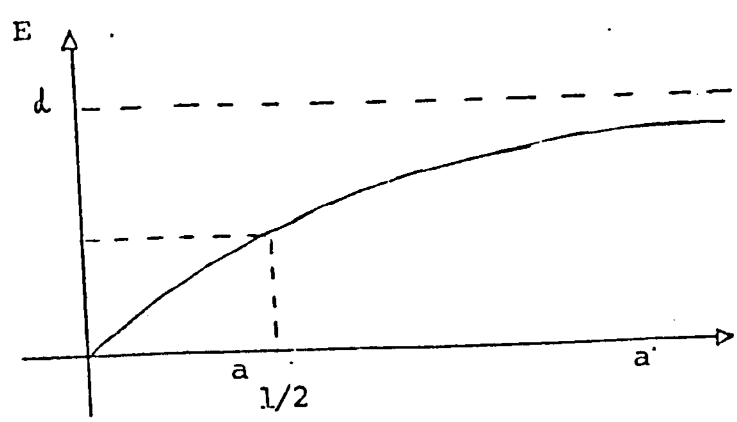
$E = d\ \{1 - \exp(-Ka)\}$, where:

E - effect of the advertising

d - is a "ceiling" factor

K - is a measure of the resistance to advertising

a - the amount of advertising which is effective.



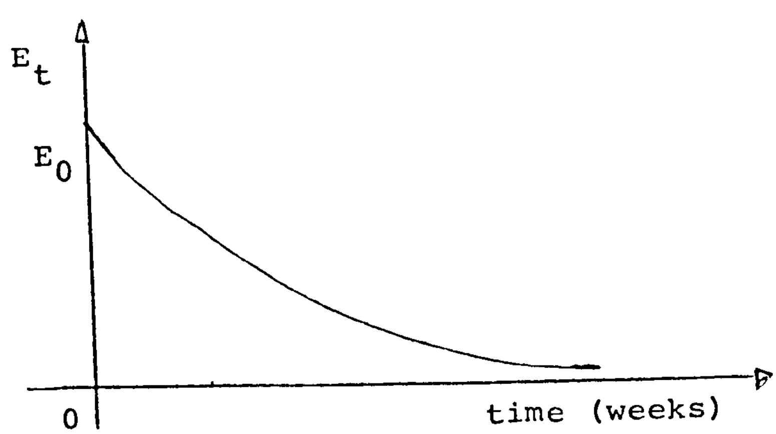
It is worth mentioning that this is the shape of the response function advocated by Simon and Arndt (1980) in their review of studies in the area.

(ii) The decay of awareness.

Let E_t be the effective awareness derived from T.V. advertising at time t. Assuming that after this week, t=0, there is no further advertising, it is hypothesised that the awareness (not the adstock) decays exponentially to some threshold value, so that E_t decays to zero. The decay rate is denoted as λ, where $0 \leq \lambda \leq 1$, giving:

$$E_t = \lambda^t . E_0 , \quad \text{where:}$$

E_0 is the initial response



Some alternative ways of considering the memory decay of a population are discussed in the appendix 4.3.1 where we suppose that the memory decay of an individual randomly selected has a beta distribution.

(iii) A sampling variation which is dependent upon the sample size and also the awareness level;

(iv) All the parameters of the model have a clear physical interpretation, as follows:

a) threshold awareness, or a base line, in the absence of advertisement - μ -

b) a maximum level or asymptotic level of awareness - $\mu+d$ -

c) the exponential decay rate - λ -;

(v) A general slow change in the quantities described in (i) to (iv), and

(vi) In response to a major change in advertisement, a likelihood of a sudden marked change in the awarensss response.

The last proposition is essentially the concept of stochastic transfer response described in Chapter 6.

4.4: Dynamic Linear Models.

4.4.1: Introduction

The effect of advertising on awareness is essentially non-linear and, although linear models can capture local descriptions of the relationship, the quantities involved in the local linear descriptions need to vary according to the effective advertising level. Consequently misinterpretation is quite likely and, in estimating the effects of different advertising levels, extrapolation will be misleading.

Broadbent (1979) discusses the relationship between awareness and T.V. advertising and although he acknowledges the non-linearity, his operational methods use linear relationships and least squares. One of his main concepts is that of adstock which, at any time t, may be thought of as measuring the current effect of all the advertising effort that has been expended. Denoting the adstock at time t by a_t, the actual TVR's for week t by x_t, and a known discount factor by λ, where $0<\lambda<1$, Broadbent calculates adstock recurrently by

$$a_t = \lambda\, a_{t-1} + x_t \quad \ldots\ldots (4.4.1)$$

Hence in the absence of further advertising the adstock decreases exponentially to zero as time increases. In practice x_t may be some function of the actual TVR's for weeks t-1 and t reflecting the sampling period with its timing variation and delayed effects. He then models the expected awareness as a linear function of adstock. For example, letting Y_t be the percentage awareness of the sample for week t, α and η be unknown constants and ν_t a random error term with zero mean, the statistical model is:

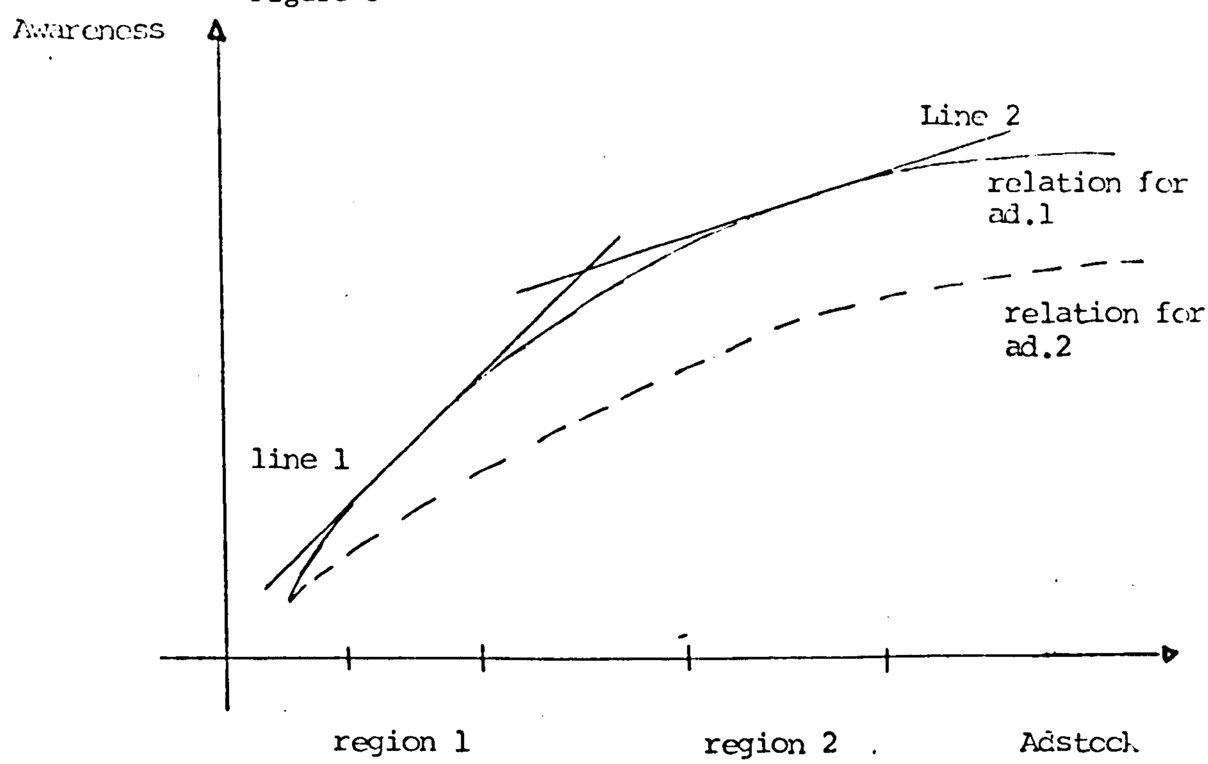
$$Y_t = \alpha + \eta a_t + \nu_t \quad \ldots\ldots (4.4.2)$$

Given historic values of the Y_t's as y_t's, λ and the x_t's the pairs (y_t, a_t) are known. The two constants (α, η) are then calculated using ordinary least squares. This model has given satisfactory results in some cases, presumably those in which the adstock did not vary greatly. However, in many others it is far from satisfactory. The effect of advertising on awareness is essentially non-linear. Although linear models can capture local descriptions, the range over which such a description is adequate is often far too small to be of practical value. This can be partly overcome by making (α, η) dynamic as in the following Dynamic Linear Models. However there are still major disadvantages in interpreting the linear model and great dangers in using such models to extrapolate to predict the outcome of proposed advertising campaigns. A number of other points also need to be made. Ordinary least squares is questionable in these applications since the random error term does not have a constant variance. Given a fixed sample size and simply arguing on a Binomially distributed response it is seen that the variance associated with an awareness level of about 50% is roughly three times as large as the variance associated with a 10% or 90% awareness level. Further the variance obviously depends upon the sample size, varying roughly in proportion to it. This indicates that sequential dynamic regression with an appropriate variance law, should improve performance. The need for λ to be known is a disadvantage. So too is the fact that a change in the televised advertisement can lead to a major sudden change in awareness.

This section considers how, by reformulating Broadbent's models, in terms of Bayesian forecasting using Dynamic Linear Models (D.L.M.'s), improvements can be obtained.

Proportional awareness varies between 0 and 1. Hence its relationship with adstock a_t can be thought of as non-linear. For example, for a given advertisement, a generally acceptable form of the relationship is as shown in Fig. 3.

Figure 3



Linear models approximate the graph by straight lines. In Broadbent's original model a single line is proposed. Its descriptive and predictive powers are obviously limited. With a Bayesian D.L.M. the appropriate line is continually updated so that in different regions of adstock, different lines are proposed. Since adstock varies relatively slowly some improvements can be obtained. When a totally different form of T.V. advertisement is used, this usually has the effect of completely altering the quantitative relationship as can be seen from Fig. 1. The Bayesian approach can be very effective in assessing such major changes.

4.4.2: The BROADYM Model.

It is a simple matter to extend Broadbent's original model to a dynamic model. Note that the original regression model of Section 2 can be written as

$$Y_t = \alpha + \eta a_t + \nu_t, \quad \nu_t \sim N[0;V]$$

$$\begin{pmatrix} \alpha \\ \eta \\ a_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} \alpha \\ \eta \\ a_{t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ x_t \end{pmatrix}, \text{ with } \lambda \text{ known}$$

The dynamic form of this model is then:

$$Y_t = \alpha_t + \eta_t \, a_t + \nu_t, \quad \nu_t \sim N[\bar{v}_t;V_t]$$

$$\begin{pmatrix} \alpha \\ \eta \\ a \end{pmatrix}_t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} \alpha \\ \eta \\ a \end{pmatrix}_{t-1} + \begin{pmatrix} w_1 \\ w_2 \\ x \end{pmatrix}_t$$

where $\begin{pmatrix} w_1 \\ w_2 \end{pmatrix}_t \sim N[\underline{\bar{w}}_t;\underline{W}_t]$.

The variance V_t may more properly reflect the sampling situation by setting $V_t = qf_t.(1-f_t)$ where $f_t = E[Y_t|D_{t-1}]$ is the expected awareness for time t based upon the information D_{t-1} up to time t-1. q is either specified as a constant, a function of the sample number or it is estimated on line. Except at times of intervention $\bar{v}_t$ and $\underline{\bar{w}}_t$ are zero. One important use of changing $(\underline{\bar{w}}_t, \underline{W})$ is when a completely new type of advert replaces an existing one. The variance $\underline{W}_t$ can be enlarged to reflect the extra uncertainty that then exists concerning the effect of the new advertisement and if required $\underline{\bar{w}}_t$ can convey the expected "improved effect" of the new advertisement as estimated by a marketing department. This is a simple example of stochastic transfer response and their estimation is dealt with in Chapter 6.

Although we have used NDLM the ideas of discount can be introduced. As an example the normal discount Bayesian Model for the Broadym may be written as

$$Y_t = \alpha_t + \eta_t a_t + \nu_t, \quad \nu_t \sim N\left[0, V_t\right]$$

Let the posterior distribution be

$$\left[\begin{pmatrix}\alpha \\ \eta\end{pmatrix}_{t-1} \Big/ D_{t-1}\right] \sim N\left[\underline{m}_{t-1};\ \underline{C}_{t-1}\right]$$

Let $0<\beta\leq 1$ be an appropriate discount factor, then the model is

$$\left[\begin{pmatrix}\alpha \\ \eta\end{pmatrix}_{t} \Big/ D_{t-1}\right] \sim N\left[\underline{m}_{t-1};\ \underline{C}_{t-1} \big/ \beta\right]$$

Defining $a_t = \lambda\ a_{t-1} + x_t$, $\underline{F} = (1,\ a_t)$ and $\underline{R} = \underline{C}_{t-1}/\beta$, the one step ahead forecast distribution is

$$\left(Y_t/x_t,\ D_{t-1}\right) \sim N\left[\hat{y}_t,\ \hat{Y}_t\right], \quad \text{where}$$

$$\begin{cases} \hat{y}_t = \underline{F}\ \underline{m}_{t-1} \\ \hat{Y}_t = \underline{F}\ \underline{R}\ \underline{F}' + V_t \end{cases}$$

Writing $\underline{A}_t = \underline{R}\ \underline{F}'/\hat{Y}_t$ and $e_t = y_t - \hat{y}_t$, where Y_t is the observed percentage awareness at time t, the necessary recurrence relationships are

$$\left[\begin{pmatrix}\alpha \\ \eta\end{pmatrix}_{t} / D_{t}\right] \sim N\left[\underline{m}_t;\ \underline{c}_t\right]$$

with

$$\begin{cases} \underline{m}_t = \underline{m}_{t-1} + \underline{A}_t\ e_t \\ \underline{C}_t = \underline{R}\left[\underline{I} - \underline{F}'\ \underline{A}'\right] \end{cases}$$

If variance learning is required let N_t be the number sampled at time t, u_t the number aware and $\hat{\underline{y}}_t$ the expected percentage of people aware. Then a variance law can be formulated as

$$V_t = q_{t-1}\ \hat{y}_t\ (100 - \hat{y}_t)/N_t$$

where $q_t = S_t/n_t$ is recurrently calculated as

$$n_t = n_{t-1} + N_t$$

$$S_t = S_{t-1} + d_t^2$$

$$d_t = (y_t - \hat{y}_t) \sqrt{N_t} \;/\; \sqrt{\hat{y}_t \, (100-\hat{y}_t)}$$

$$y_t = 100 \, u_t/N_t$$

Further details of variance learning can be found in West (1983), Ameen and Harrison (1983), and Discount Bayesian Regression is discussed in Harrison and Johnston (1983).

4.4.3: The Local Linear Model.

A weakness of the previous models is the need to pre-specify the unknown decay factor λ. It is an advantage if λ can be estimated on-line and possibly be allowed to vary with time and type of advertisement. Taking equations 4.4.1 and 4.4.2 of Broadbent's original model, given in Section 4.4.1, it is easily seen that

$$y_t - \lambda \, y_{t-1} = \mu + \eta \, x_t + \nu_t - \lambda \, \nu_{t-1} \qquad (4.4.3)$$

where $\mu = (1-\lambda)\alpha$. This leads to the proposal of the DLM $\{\underline{F}_t, I, V_t, \underline{W}_t\}$ where $\underline{F}_t = (1, x_t, y_{t-1}, e_{t-1})$ and $\underline{\theta}_t' = (\mu,\eta,\lambda,\gamma)_t$. The model may be written in full as:

$$y_t = \mu_t + \eta_t \, x_t + \lambda_t \, y_{t-1} + \gamma_t \, e_{t-1} + \nu_t$$

$$\begin{pmatrix} \mu \\ \eta \\ \lambda \\ \gamma \end{pmatrix}_t = \begin{pmatrix} \mu \\ \eta \\ \lambda \\ \gamma \end{pmatrix}_{t-1} + \underline{w}_t \qquad (4.4.4)$$

where e_t is the usual one step ahead forecast error, $\nu_t \sim N(\bar{v}_t; V_t)$ and $\underline{w}_t \sim N(\underline{\bar{w}}_t; \underline{W}_t)$. Generally $\underline{W}_t$ is a small variance matrix allowing the parameter vector $\underline{\theta}$ to vary through time. However, when exceptional events occur, such as changes in the form of advertisement, the appropriate elements of $\underline{W}_t$ are enlarged to provide a rapid means of estimating the effect of the event.

4.4.4: Comments

The D.L.M.'s can adapt well to slow changes in the intercept and slope of the local line but obviously, from Fig. 1, their extrapolative validity is unsatisfactory, as is their global description of the true relationship. Nevertheless for short-term description and short term retrospection, the model will be reasonable for the greater part of the time. Further the models are designed to deal quickly with sudden major changes in awareness, which for example, may occur with a qualitative change in advertising. Referring to Fig. 3, if advert 1 is to be replaced by advert 2 whose awareness response form is unknown, a signal of the new advertisement can raise the uncertainty associated with this response and provide a very fast means of adaptation. Even so, comparisons between adverts is not easy. If linear characteristics are used, the estimated difference may be confounded by differing amounts of expenditure on the two adverts. Missing observations cannot be properly handled by many classical time series models but, as it is well known, Bayesian methods have no difficulty with them. Further, observed awareness is more properly modelled based on the Binomial distribution. The varying variance V_t can be designed to reflect this and provided the sample is sufficiently large a Normal distribution can be assumed for Y_t. However, for smaller samples and either low or high awareness levels it may be desired that a Binomial Distribution is used. An example of such a linear model is given as a particular case of the non-linear models in section 5.3. Further that example demonstrates how the extra variation due to sampling in what might often be an unrepresentative T.V. region can be modelled.

4.5: Dynamic Non-Linear Models.

4.5.1: General.

In this particular application the non-linear models have been developed based on the propositions described in Section 4.3. Some consequences of the basic hypothesis are derived and the Dynamic Non-Linear Models of Chapter 3 are applied.

4.5.2: The Assumed Relationship Between Awareness and Advertising.

The relationship between awareness and advertising is derived as a consequence of the hypothesis of diminishing return and the exponential decay of awareness in the absence of further advertising (Section 4.3). An interesting relation between adstock at time t-1 and at time t is also concluded.

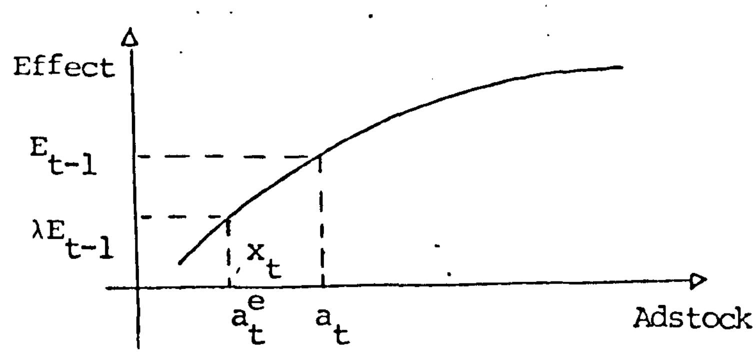
(i) In the absence of an advertising input in week t, the adstock reduces from a_{t-1} to a_t as a consequence of relations 4.3.2 (i) and (ii). These quantities are related according to:

$$a_t = \frac{-1}{k_t} \log \left(\frac{d_t - E_t}{d_t} \right)$$

$$= \frac{-1}{k_t} \log \left[1 - \lambda_t \{ 1 - \exp (- k_t a_{t-1}) \} \right]$$

Hence, in these models, adstock does not reduce exponentially with time: it is the memory effect of adstock that decays exponentially.

(ii) Let E_{t-1} be the effect of advertising at time t-1, so from 4.3.2 (ii) this effect will decay to λE_{t-1} at time t. The equivalent adstock at time t-1 is obtained from (i) above, as:



$$a^e_{t-1} = - \frac{1}{k_t} \log \left(\frac{d_t - \lambda_t E_{t-1}}{d_t} \right) \tag{4.5.1}$$

In this way the adstock at time t is:

$$a_t = a^e_{t-1} + x_t$$

where x_t is the advertising input at time t.

The final relationship between advertising and awareness effect is:

$$E_t = d_t - \left[d_t - \lambda_t E_{t-1}\right] \exp(-k_t x_t) \qquad (4.5.2)$$

4.5.3: Notation and Observational Model.

t	denotes week t;
μ_t	the threshold awareness or base level, which awareness falls to when there is no ad-stock.
$\mu_t + d_t$	the asymptotic awareness;
λ_t	the memory decay rate;
x_t	the number of TVR's in week t (possibly delayed);
N_t	the number of people sampled in week t;
u_t	the number of people "aware" in week t;
ρ_t	the proportional awareness in the population;
$Y_t = \frac{100u_t}{N_t}$	the observed percentage awareness in week t;
E_t	the effect of TV advertising up to and including time t on the awareness at time t;
k_t	the advertising response factor.

The Population Awareness

The population awareness p_t is then modelled using a derived guiding relationship

$$p_t \simeq \mu_t + d_t - \left[d_t - \lambda_t E_{t-1}\right] \exp(-k_t x_t) \qquad (4.5.3)$$

Since the weekly awareness sample is taken in selected T.V. regions the relevant regional awareness ψ_t will vary around p_t and is modelled as

$$\psi_t \sim \left[p_t;\ \sigma^2\right].$$

This can be expressed as $\psi_t = p_t + \delta$ with $\delta \sim \left[0;\ \sigma^2\right]$.

The Observed Awareness

Letting N_t people be sampled in week t, the number of aware people U_t is modelled as a conditional Binomial

$$(U_t | \psi_t, N_t) \sim \text{Binomial } [N_t, \psi_t].$$

If N_t is sufficiently large, as is often the case, both U_t and the percentage awareness Y_t can be well approximated by a Normal Distribution with

$$(Y_t | \psi_t, N_t) \sim \left[100\psi_t/N_t ; 10^4 \ \psi_t \ (1-\psi_t)/N_t\right]$$

Measuring the Effect of an Advertisement.

One measure of a given advertisement is $T = (\log_e 2)/K = 0.7/k$. This may be regarded as a measure of the resistance to advertising. The interpretation is that T is the number of T.V.R.'s required to achieve a half of the possible remaining potential awareness which at any time t is $d_t - E_t$. This measure allows a comparison between different advertisements.

In one of the later models, K is fixed for all advertisements and then different advertising campaigns may be compared using the quantities $\mu_t + d_t$. Further discussion of measures used in practice is to be found in Colman and Brown (1983).

4.5.4: Non-Linear Model with Variable Half-Effects.

The non-linear model is completely specified by a set of guide relationships and the observational distribution. With the normality assumption for the marginal distribution for the underlying parameters $\underline{\theta}_t$ there is no technical problem since the first two moments of the guide relationship can be calculated.

In this model the value of the asymptote $\mu_t + d_t$ = constant is to be specified, although μ_t and d_t are allowed to vary along the time.

<u>Guide Relationships</u>: Following 4.5.3 the guide relationship for the regional proportional awareness ψ_t is to be taken as

$$g(\underline{\theta}_t) \simeq \mu_t + d_t - [d_t - \lambda_t \ E_{t-1}] \exp[-k_t \ x_t] + \delta_t$$

where: $\underline{\theta}_t = (\mu_t, d_t, \lambda_t, k_t, E_{t-1})'$

$\delta_t \sim [0, \sigma^2]$

The corresponding guide relationship for E_t is

$$E_t = d_t - (d_t - \lambda_t E_{t-1}) \exp[-k_t x_t]$$

However, for computational ease, in this model, a Taylor series expansion is used just in updating, so that we use:

$$E_t \simeq \lambda_t E_{t-1} + (d_t - \lambda_t E_{t-1}) k_t x_t \qquad (4.5.4)$$

and: $g(\underline{\theta}_t) = \mu_t + E_t + \delta_t$

Systems Evolution: Let the information up to and including time t-1 be

$$(\underline{\theta}_{t-1}/D_{t-1}) \sim N [\underline{m}_{t-1}, \underline{C}_{t-1}], \text{ with } \underline{C}_{t-1}$$

constrained by $c_{11} = c_{22} = c$, $c_{12} = c_{21} = -c$, to ensure that $\mu+d$ is fixed equal a constant, and suppose that the link between (θ_t/D_{t-1}) and $(\underline{\theta}_{t-1}/D_{t-1})$ is:

$$\underline{\theta}_t = \begin{bmatrix} I_4 & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & 1 \end{bmatrix} \begin{bmatrix} \underline{\theta}_{t-1} \\ E_{t-1} \end{bmatrix} + \underline{\omega}_t$$

with $\underline{\omega}_t \sim N \left[\underline{\bar{\omega}}_t, \underline{W}_t\right]$, where $w_{11} = w_{22}$ w and $w_{12} = w_{21} = -w$

The distribution of $\left[\begin{matrix} E_t \\ \underline{\theta}_t \end{matrix} \middle/ D_{t-1}\right]$ is then derived using the above guide relationship for $\underline{E}_t$; and $\psi_t = g(\underline{\theta}_t)$ provides

$$\left(\begin{matrix} \psi_t \\ E_t \\ \underline{\theta}_t \end{matrix} \middle/ D_{t-1}\right) \sim \left[\begin{pmatrix} \hat{\psi} \\ \hat{E}_t \\ \hat{\theta}_t \end{pmatrix} ; \begin{bmatrix} r & \underline{r} \\ \underline{r}' & \underline{R} \end{bmatrix}\right]$$

Observational Model: the observation model may be specified as Binomial or as a Normal using a variance governed by the Binomial characterization. Following the principles described in the sections dealing with linear models all the request facilities as given in 4.2 are provided.

Forecast Function: In obtaining the forecast function which provides future point estimates of consumer awareness the full non-linear form

$$g(\hat{\theta}_t) = \hat{\mu}_t + \hat{d}_t - \left[\hat{d}_t - \hat{\lambda}_t \hat{E}_{t-1}\right] \exp(-\hat{k}_t x_t)$$

is used with modal estimates $\hat{\underline{\theta}}_t$ being used for all future point estimates. Note that this is one application of the global form forecasting function as discussed in Harrison-Leonard-Gazard (1977).

4.5.5: Non-Linear Model with Fixed Half-Effect.

Guide Relationship: In this model the guide relationships are defined as:

$$E_t \simeq d_t \{1-\beta(x_t)\} + \lambda_t E_{t-1} \beta(x_t) \qquad 4.5.5$$

where

$$\beta(x_t) = \exp\left[-k_0 x_t\right], \; k_0 \text{ - fixed; and}$$

$$g(\underline{\theta}_t) = \mu_t + E_t + \delta_t; \; \delta_t \sim [0,\sigma^2] \text{ as before.}$$

System evolution: The evolution is described by the relation

$$\underline{\theta}_t = \underline{G}\, \underline{\theta}_{t-1} + \underline{\omega}_t$$

where:

$$\underline{G} = \begin{bmatrix} \underline{I}_4 & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & 1 \end{bmatrix}$$

$$\underline{\omega}_t \sim N\left[\bar{\underline{\omega}}_t; \underline{W}_t\right]$$

So the prior distribution is obtained using the guide relationships as:

$$\left(\begin{matrix} \psi_t \\ E_t \\ \underline{\theta}_t \end{matrix} \Big/ D_{t-1} \right) \sim \left[\begin{pmatrix} \hat{\psi} \\ \hat{E}_t \\ \hat{\theta}_t \end{pmatrix} ; \begin{bmatrix} r & r \\ \underline{r}' & \underline{R}_t \end{bmatrix} \right]$$

Observational Model: The observational model may be specified as normal with the following variance-law

$$V_t = q \cdot \hat{y}_t(100 - \hat{y}_t)/N_t$$

Finally it is worth commenting that the updating procedure is as described in Chapter 2 for the non-linear normal models and, if desired, the variance learning discussed in Section 4.4.2 can be used.

4.6 Examples and Comparison.

4.6.1: General

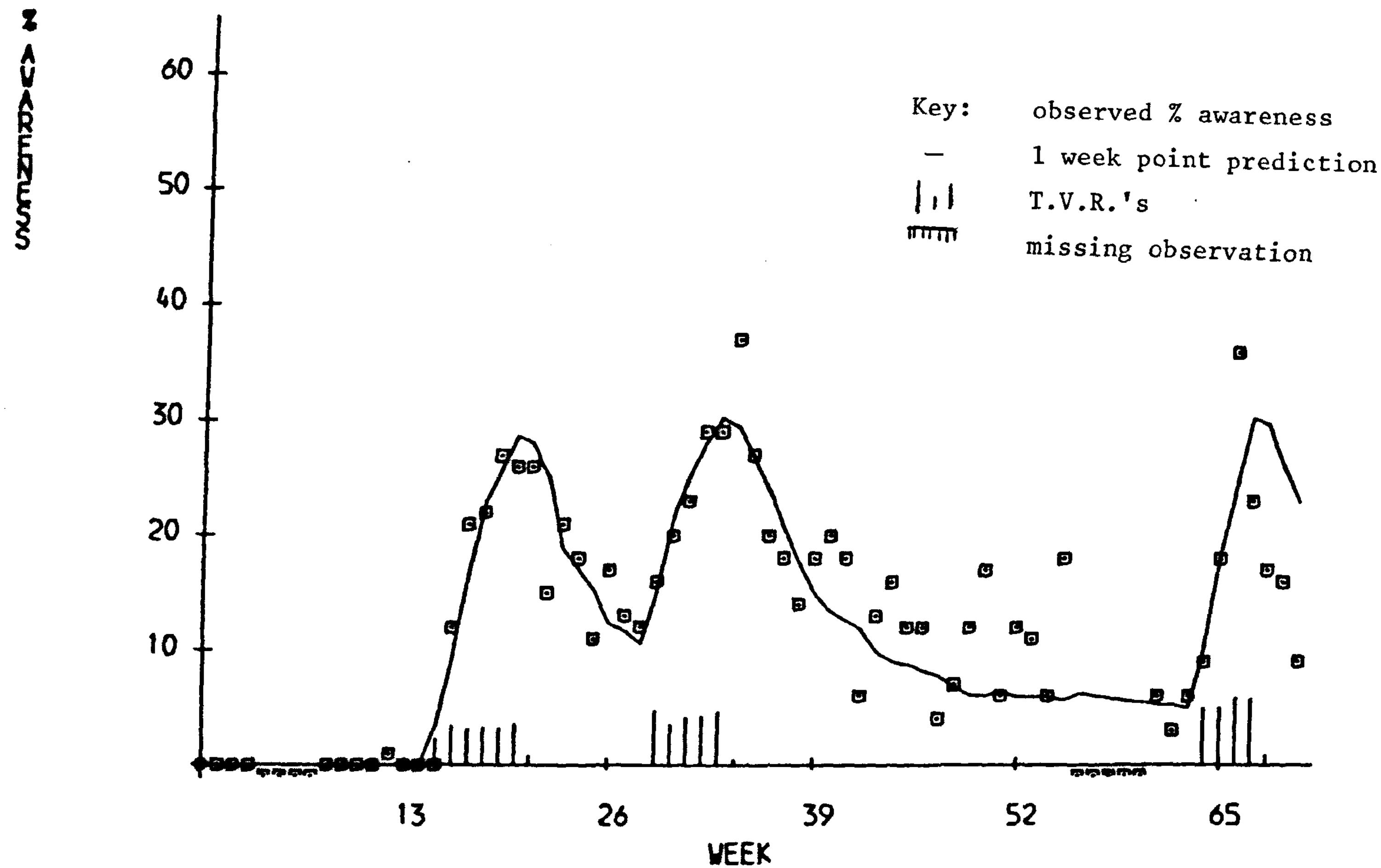
A number of examples of the application of the models are given in Colman and Brown (1983) where they are discussed from a practitioners viewpoint. Estimates of some of the parameters as they evolve throughout time and of the effectiveness of different types of advertisement may be found in that paper. In this thesis we concentrate on comparing the relative performance of the linear and non-linear models. It is shown that, over a sample of products, there is little to choose between the two linear models and that there is also little difference between the two non-linear models. However, the predictive performance and the descriptive power of the non-linear models is seen to be far better than that of the linear models.

The initial setting of the parameters and the posterior mean and standard deviation for each data set and each model can be found in the appendix 4.2 and 4.3 respectively.

4.6.2: Examples

In the following examples we present the data for the main products involved in this evaluation study. The performance of the models is assessed on a qualitative basis, that is to say emphasising the descriptive power of the models.

The observed weekly percentage consumer awareness and the weekly TVR's on product P1 are plotted for a period of 70 weeks. The one week ahead point


% AWARENESS
60
50
40
30
20
10
13
26
39
52
65
WEEK
Key:
observed % awareness
1 week point prediction
T.V.R.'s
missing observation

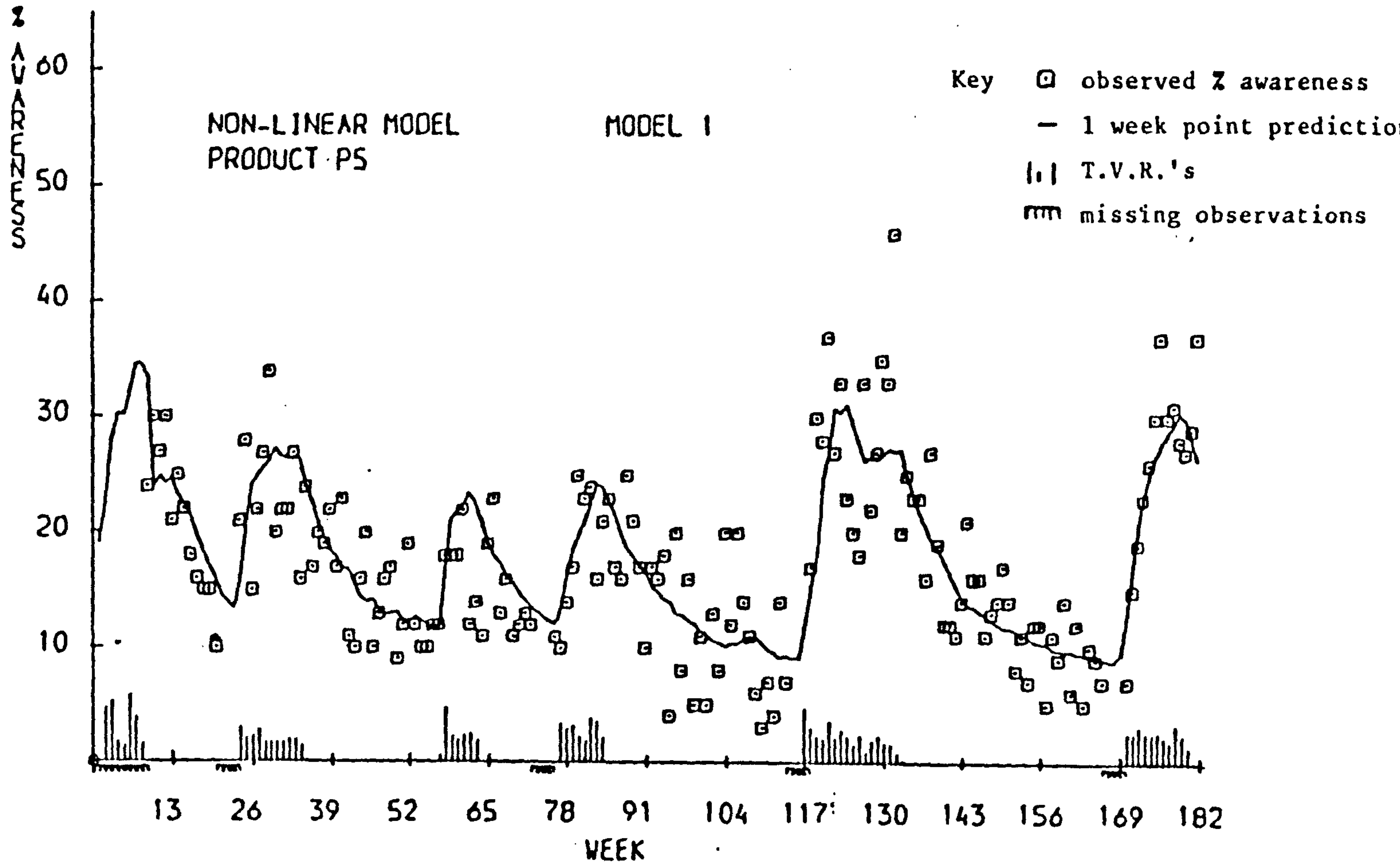
predictions using model 2 (the non-linear model with half-effect fixed) are also plotted. The non-linear relationship between TVR and awareness can be appreciated and the decay in awareness when no further advertising is coming in is evident.

In Fig. 4.6.2 the data on product P5 is given. This is a fast moving consumer good. Both the observed weekly percentage consumer awareness and the weekly TVR's are plotted for a period of more than three years. Periods of missing observations, which correspond to no awareness samples, are indicated. Plotted also are the one week ahead point predictions derived using Model 1. The decay in awareness during periods of no advertising is evident on a number of occasions and the sharp non linear response to an advertising burst can also be appreciated. For this product, there are five periods of missing observations.

A comparison between the Local Linear model and the non-linear model 1 is shown in Fig. 4.6.3 using data on Product P2. The point predictions are for four weeks ahead. Fig. 4.6.4 shows a similar comparison for the linear model Broadym and the non-linear model 2. The data here relates to a chocolate bar which is also examined in Colman and Brown (1983) whose later data is included. The interesting feature of this product is the use of three distinct advertising campaigns. The first advertisement was successful in generating consumer awareness. However in weeks it was replaced by a qualitatively different advertisement which unfortunately was soon assessed by the models to be relatively poor. In fact this latter advertisement was estimated to require almost 5 times as many TVR's as the former advertisement in order to achieve the same increase in awareness response. In turn this advertisement was replaced in week 158 by another qualitatively different advertisement that was estimated to improve the awareness response over its immediate predecessor by a factor of about 3.

Figures 4.6.5 and 4.6.6 show the extrapolative inadequacy of the linear models. In Fig. 4.6.5 product P2 is shown with the point predictions made at week 39 for the future advertising programme over the next 31 weeks. Clearly compared to the performance of the non-linear model, the linear model fails to give a satisfactory prediction.

FIG. 4.6.2



% AWARENESS
60
50
40
30
20
10
NON-LINEAR MODEL
PRODUCT PS
MODEL 1
Key
observed % awareness
1 week point prediction
T.V.R.'s
missing observations
13
26
39
52
65
78
91
104
117
130
143
156
169
182
WEEK

FIG. 4.6.3

Key: as for fig.4.6.2

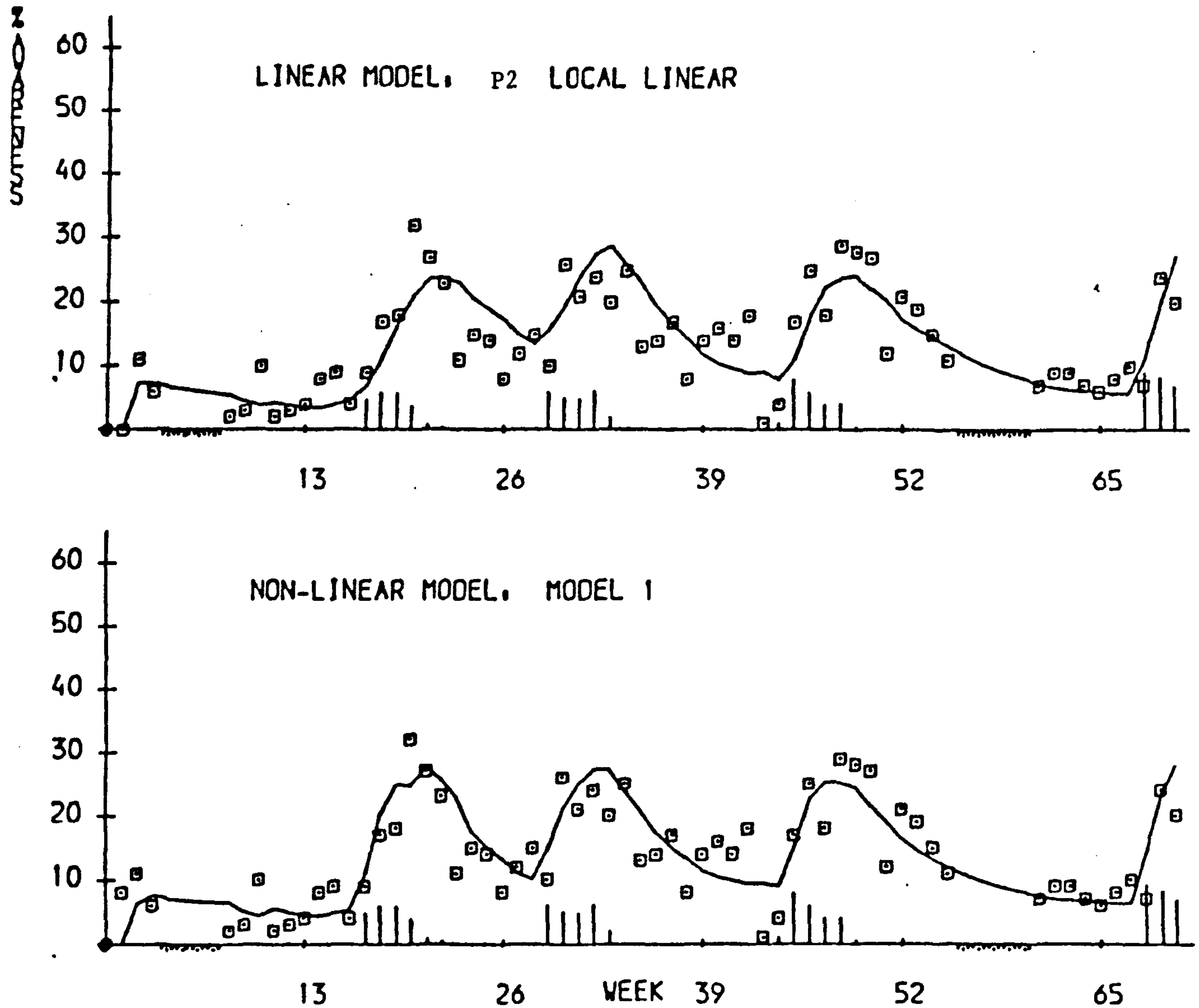


FIG. 4.6.4

Key: as for Fig.4.6.3 , except

- four week ahead point prediction.

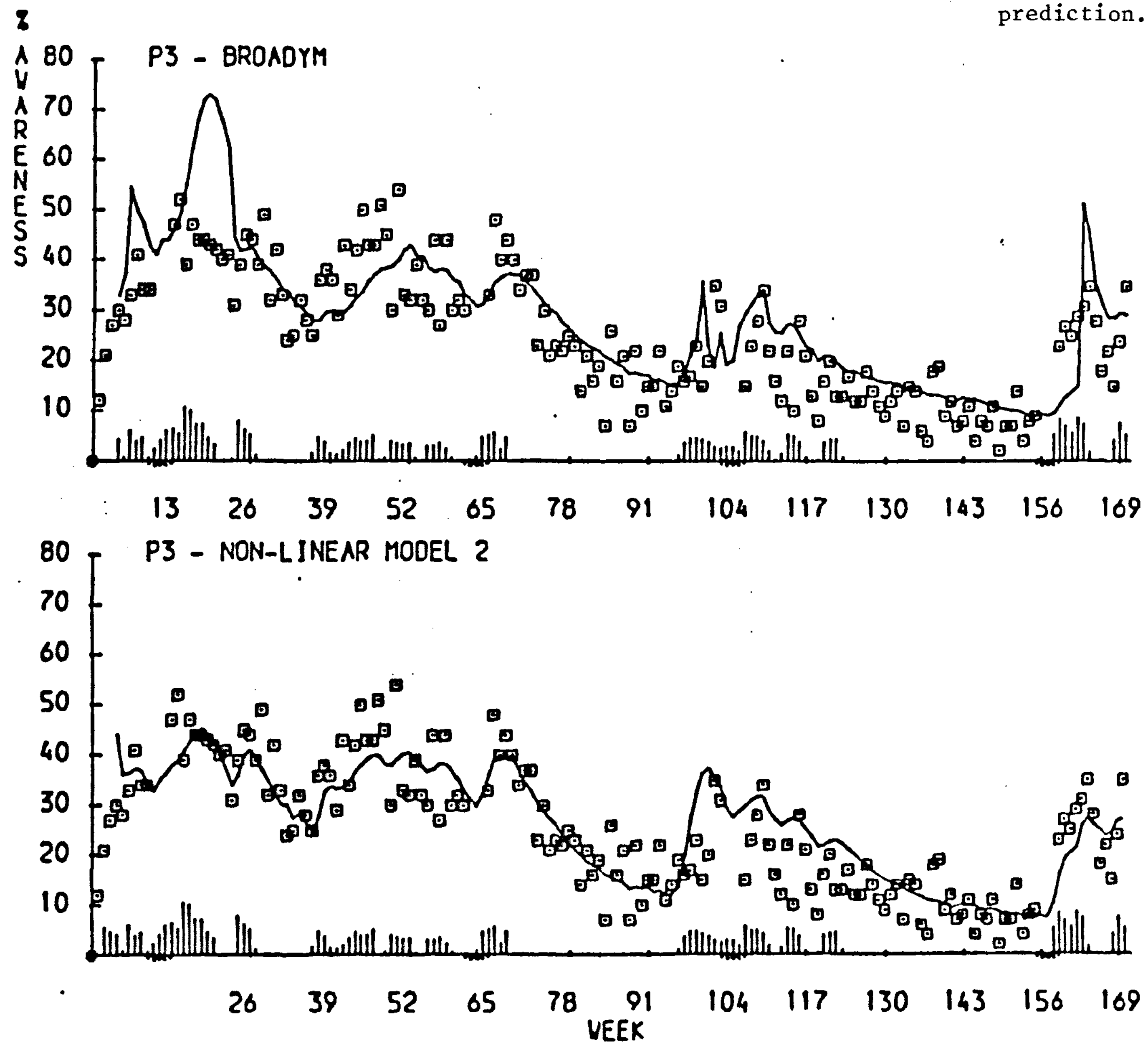
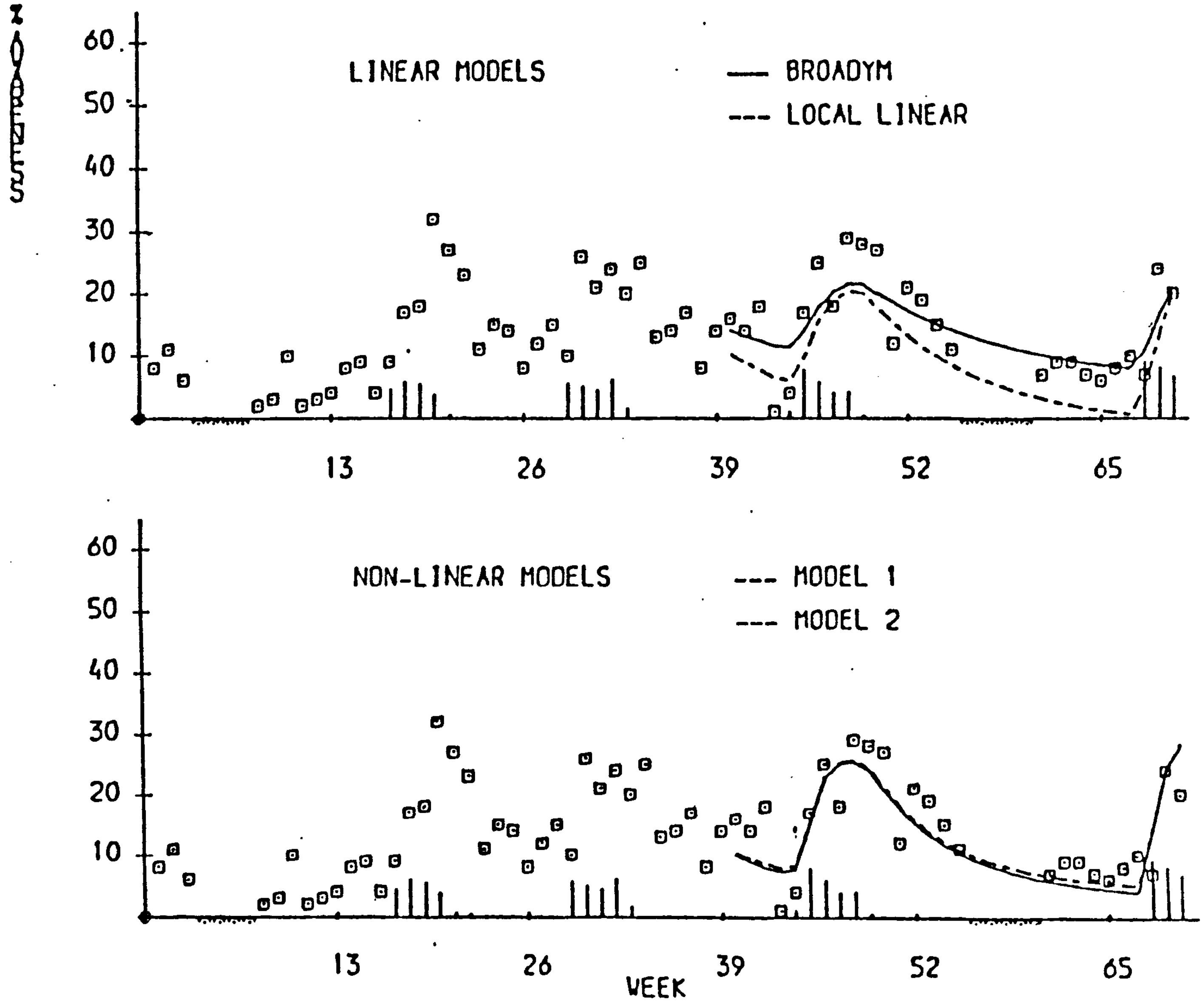


FIG. 4.6.5 : Product P2

Point Prediction made at week 39 for the next 31 weeks.



Similarly in Fig. 4.6.6 with product P5, the point predictions made at week 129 for the future advertising programme over the next year of 53 weeks is given. Again the performance of the linear model is totally unsatisfactory and shows that for "what if" analyses the linear models are not to be recommended.

Finally in Fig. 4.6.7 we have the data on product P1 and the extrapolative performance of the models. For this particular data set the four models are quite similar.

4.6.3: Comparison

Tables 1 and 2 give a comparison of the four models relative to that of the non-linear model 1. The criteria is the inverted normal loss function, which was used in order to avoid the effect of some outliers.

As we see from tables a and b, in the appendix 4.4, the two criteria, when expressed as integers, give identical results. In the appendix 4.3 we present some additional results using the point prediction for the future advertising programmes.

The performance for four steps ahead using non-linear model 1 was assessed without signalling the change in the quality of advertisement by increasing the uncertainty of the relevant parameters. The loss was 32% greater showing the large benefits to be gained by considering the response to different advertisements as stochastic.

FIG. 4.6.6 : Product P5

Point Prediction made at week 129 for the next 53 weeks.

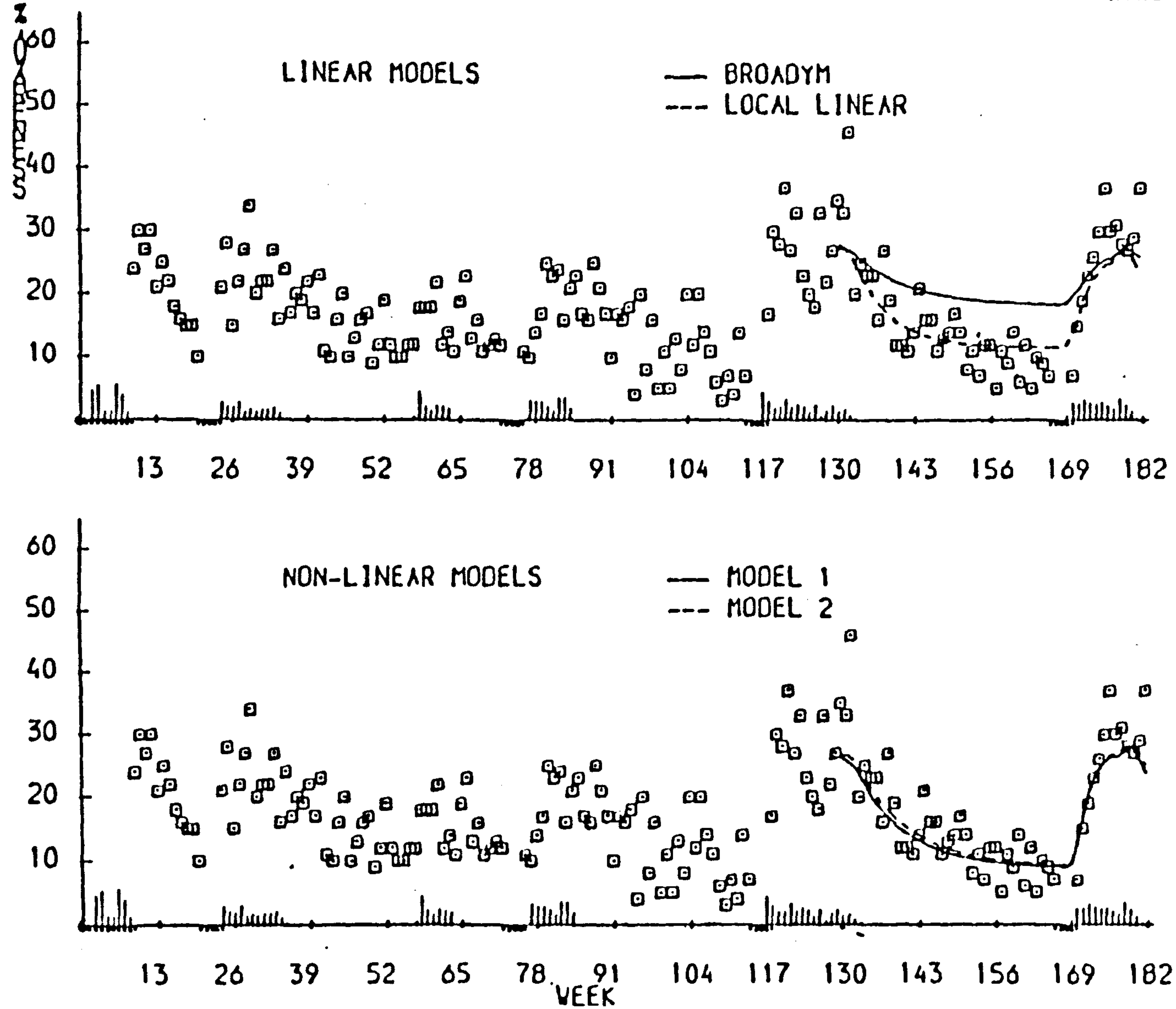


FIG. 4.6.7

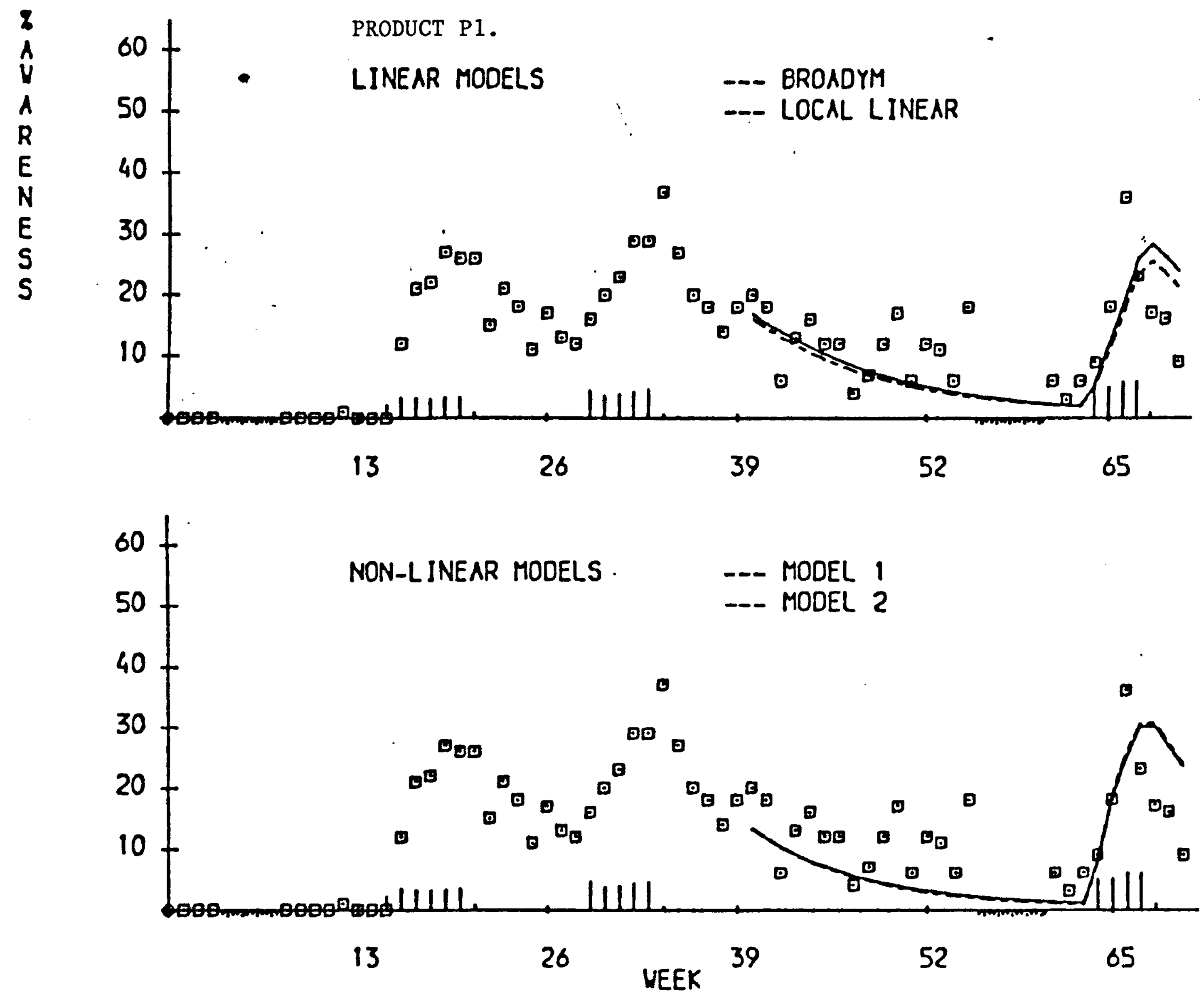

PRODUCT P1.
LINEAR MODELS
--- BROADYM
--- LOCAL LINEAR
% AWARENESS
60
50
40
30
20
10
13
26
39
52
65
NON-LINEAR MODELS
--- MODEL 1
--- MODEL 2
60
50
40
30
20
10
13
26
39
52
65
WEEK

TABLE 1. A COMPARISON OF THE FOUR MODELS.

The tabulated entries are the percentage loss in excess of that incurred by model 1, using an inverted Normal loss function.

One week ahead point forecast performance

PRODUCT	BROADYM	LOCAL LINEAR	MODEL 2	MODEL 1
P1	35	47	2	0
P2	21	26	5	0
P3	18	9	1	0
P4	2	-1	-5	0
P5	8	20	6	0
MEAN	17	20	2	0

Four week ahead point forecast performance

PRODUCT	BROADYM	LOCAL LINEAR	MODEL 2	MODEL 1
P1	81	80	0	0
P2	31	41	11	0
P3	56	15	-7	0
P4	4	21	-9	0
P5	31	32	3	0
MEAN	41	38	0	0

Appendix 4.1 Memory Decay of a Population

We suppose that the memory decay λ of an individual randomly selected is distributed over (0,1) as a Beta, so that

$$f(\lambda) = \frac{1}{B(a,b)} \lambda^{a-1} (1-\lambda)^{b-1}; \ a>0, \ b>0$$

It follows that

$$\text{Mode}(\lambda) = \frac{a-1}{a+b-2} \quad \text{or } 0 \text{ or } 1$$

$$E(\lambda) = \frac{a}{a+b}$$

$$v(\lambda) = \frac{ab}{(a+b+1)(a+b)} \quad \text{and}$$

$$E(\lambda^k) = \frac{\Gamma(k+a) \ . \ \Gamma(a+b)}{\Gamma(a) \ . \ \Gamma(k+a+b)}$$

The case b=1 is worth considering since the mode is at 1. In this case the expectation of λ^k reduces to

$$E(\lambda^k) = \frac{a}{k+a}$$

which shows that the memory decay for the population is slightly bigger than a simple exponential decay as we can see in the following table

K	λ=,9	$E\lambda^k$; a=9
0	1	1
1	.9	.9
2	.81	.818
3	.73	.75
4	.59	.69

Appendix 4.2: The Initial Setting for Parameters.

This appendix contains the initial setting of the parameters for the various products and models.

BROADYM:

	MEAN	VARIANCE	DISTURBANCE
level	10.0	100.00	.04
slope	.10	.04	.04 E-4
λ	.90	-	-
a_0	.00	-	-

except for products P_1 and P_2 where the level mean are: 0.0

LOCAL LINEAR:

	MEAN	VARIANCE	DISTURBANCE
level	4.00	4.00	0.1
slope	.04	.0004	.01 E-4
λ	.90	.0025	.01 E-4
γ	- .80	.0025	.01 E-4
a_0	200.0		

except for products P_1 and P_2 where $\overline{\mu} = 0.0$ and $\overline{E} = 0.0$

MODEL 2

	MEAN	VARIANCE	DISTURBANCE
level	16.00	64.00	.04
D	44.00	64.00	.04
λ	.90	.0064	.01 E-4
E	16.00	64.00	.00
H	200.00		

except for P1 and P2 where $\bar{\mu}$ = 0.0 $\bar{E}$ = 0.0 and H = 1/k.

MODEL 1

	MEAN	VARIANCE	DISTURBANCE
μ	0.00	64.00	.04
D	45.00	64.00	.04
λ	.90	.0064	.01 E-4
H	200.00	1000.00	2.40
E	0.00	64.00	0.00

except for products P_3 $\bar{\mu} = 10$ $\bar{d} = 35$

P_4 $\bar{\mu} = 30$ $\bar{d} = 55$

P_5 $\bar{\mu} = 10$ $\bar{d} = 35$

Appendix 4.3

For each product and each model we present below the posterior mean and standard deviation at the end of the data set.

BROADYM

with λ = .90 fixed

PRODUCT	LEVEL MEAN	LEVEL SD	SLOPE MEAN	SLOPE SD
P1	3.32	1.54	.058	.01
P2	4.52	1.73	.043	.01
P3	13.32	6.32	.024	.01
P4	42.53	2.60	.036	.01
P5	8.52	1.69	.074	.01

LOCAL LINEAR

	LEVEL		SLOPE		DECAY		GAMMA	
	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
P1	.29	1.03	.07	.02	.88	1.04	-.75	.05
P2	.58	1.07	.06	.01	.87	.05	-.79	.05
P3	.41	1.90	.06	.02	.84	.04	-.76	.05
P4	6.35	1.75	.08	.02	.84	.03	-.62	.04
P5	1.45	1.21	.07	.02	.87	.04	-.76	.05

MODEL 1

	μ		d		H		λ	
	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
P1	3.62	2.01	41.38	2.01	261.19	36.8	.89	.02
P2	4.82	2.05	40.18	2.05	201.67	49.4	.84	.04
P3	9.87	1.53	55.13	1.53	743.17	134.4	.86	.02
P4	31.37	4.07	53.63	4.07	494.40	114.9	.95	.01
P5	7.99	1.70	37.01	1.70	210.17	23.9	.90	.02

MODEL 2

	μ		d		λ	
	MEAN	SD	MEAN	SD	MEAN	SD
P1	3.25	2.12	35.48	3.80	.89	.02
P2	4.05	2.19	35.75	3.81	.85	.04
P3	6.92	1.99	26.08	3.44	.91	.02
P4	23.93	4.93	41.72	6.20	.96	.01
P5	7.98	1.74	37.00	3.58	.90	.02

Appendix 4.4: A Comparison of the Four Models.

The tabulated entries are the percentage loss in excess of that incurred by model 1, using an inverted Normal and quadratic loss function.

Table a: Inverted Normal Loss Function

One week ahead point forecast performance

PRODUCT	BROADYM	LOCAL LINEAR	MODEL 2	MODEL 1
P1	34.79	46.87	2.09	0
P2	20.69	25.90	4.63	0
P3	17.75	9.23	.42	0
P4	2.09	-1.06	-5.38	0
P5	8.48	20.12	6.27	0

Four week ahead point forecast performance

PRODUCT	BROADYM	LOCAL LINEAR	MODEL 2	MODEL 1
P1	81.20	79.85	- .29	0
P2	31.43	41.02	11.17	0
P3	55.94	14.88	-6.91	0
P4	4.15	20.85	-9.15	0
P5	31.31	31.79	3.23	0

TABLE b - QUADRATIC LOSS FUNCTION

One Week ahead

PRODUCT	BROADYM	LOCAL LINEAR	MODEL 2	MODEL 1
P1	34.83	36.90	2.10	0
P2	20.72	25.95	4.63	0
P3	17.78	9.24	.91	0
P4	2.09	-1.07	-5.39	0
P5	8.49	20.14	6.29	0

Four Week ahead

PRODUCT	BROADYM	LOCAL LINEAR	MODEL 2	MODEL 1
P1	81.43	79.99	-.28	0
P2	31.48	41.08	11.18	0
P3	56.17	14.88	-6.90	0
P4	4.16	21.08	-9.34	0
P5	31.43	31.82	3.22	0

Chapter 5. A general Non-Linear Model.

5.1: Introduction.

The class of dynamic non-linear normal model developed in Chapter 3 is extended to the non-normal case. The approach is based on dynamic extention of non-linear regression models in the one parameter exponential family of distributions, the class of generalized linear models.

Conjugate prior and posterior distributions for the exponential family are used leading to the calculation of predictive distributions for forecasting in closed standard form.

A review of the static GLM is introduced, emphasizing the structure of the models, as a base for the extention to the non-normal case. In Section 5.3 we discuss the dynamic generalized linear model and we present the method of feedback of observational information to the state parameters. The linear Bayes prediction method is used.

Finally in Section 5.4 the class of general dynamic non-linear model is developed and some examples are discussed.

5.2: A review of the static generalized linear model.

5.2.1: General.

A general approach to the analysis of static regression problems in the exponential family was given by Nelder and Wedderburn [1972]. The generalized linear model, as they called it, comprises a wide class of models suitable for application in many practical problems and the methodology has spread rapidly due to the availability of comprehensive computer packages (Baker and Nelder [1978]). A Bayesian inference and data analysis for the generalized linear model is discussed in West [1983].

The class of generalized linear models includes techniques developed for non-normal data such as probit analysis, where a binomial variate has a parameter related to an assumed underlying distribution; contingence tables, where the distribution is multinomial or Poisson with constraints and the systematic part of the model usually multiplicative; and estimation of variance component from independent quadratic forms, where the distribution and the systematic component has a linear structure.

In all the above mentioned examples the following characteristics are present:

(i) a dependent variable, say Y, whose distribution, with parameter ψ, is a member of the exponential family;

(ii) a set of independent variables, $x_1, x_2 \dots x_p$, and a linear systematic component:

$$\lambda = \sum_{i=1}^{p} \theta_i \, x_i \quad ,$$

(iii) a linking function $g(\psi)$ connecting the parameter ψ of the distribution of Y with the λ's of the linear model.

In the remainder of this section we briefly discuss the linear model for the systematic effects, and some properties of the exponential family as a useful class of distributions to represent the random part of the model.

5.2.1: The linear model for sytematic effects.

Suppose that $x_1, \dots, x_p$ are p-independent variables whose values are known and denote by θ_i, i=1,2,...p, a set of p unknown parameters whose estimate is required.

The linear model is: $\lambda = \sum_{i=1}^{p} \theta_i \, x_i$, where the x variates may be qualitative or quantitative.

5.2.2: The error structure.

In order to represent the random component of the GLM a general class of distributions is introduced. Suppose that the observation Y comes from a distribution with density function given by:

$$p(y/\psi,\phi) = \exp\left[\phi\{y\psi+a(\psi)\}\right] b(y,\phi) \qquad 5.2.1$$

where $\phi>0$ and ψ is the natural parameter of the distribution.

The class of the densities of the type 5.2.1 will be called the general exponential family, and the mean and variance of Y can be expressed in terms of ϕ and ψ as:

(i) $E[Y/\psi,\phi] = \mu = a'(\psi)$

(ii) $V[Y/\psi,\phi] = a''(\psi)/\phi$, where ' means the derivative of a.

The results in (i) and (ii) can be easily proved using the identities (see Cox and Hinkley (1979); pp. 107-108).

$$E\left[\frac{\partial}{\partial\psi} \lg p(y/\psi,p)\right] = 0$$

$$E\left[\frac{\partial^2}{\partial\psi^2} \lg p(y/\psi,\phi)\right] = - E\left[\frac{\partial}{\partial\psi} \lg p(y/\psi,\phi)\right]^2$$

where some regularity conditions are assumed to guarantee the interchangeability of the order of integration and differentiation.

Then it follows that:

(i) $E\left|\frac{\partial}{\partial\psi}\{\phi[Y\psi-a(\psi)] + \lg b(y,\phi)\}\right| =$

$= E[\phi\{Y-a'(\psi)\}] = 0;$ and so:

$\mu = E[Y/\psi,\phi] = a'(\psi)$

(ii) $-E[\phi a''(\psi)] = -\phi a''(\psi)$

$= -E\{\phi[Y-a'(\psi)]\}^2$

$= - \phi^2 v(Y/\psi,\phi);$ so:

$v(Y/\psi,\phi) = a''(\psi)/\phi.$

5.2.3: Notation and Examples.

The notation for some special cases of interest is as follows:

(a) Normal Distribution

Here we have:

$$p(Y/\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2} \{Y-\mu\}^2\right], \qquad x \in \mathbb{R}$$

$$= \exp[\phi\{Y\mu-\mu^2/2\}].b(Y,\phi), \text{ where:}$$

$\phi = 1/\sigma^2$, $\psi=\mu$ and $a(\theta) = \psi^2/2$; and

$E(Y/\psi.\phi) = \mu$ and $v(Y/\psi,\phi) = \sigma^2$

(b) Gamma Distribution.

Let $(Y/\theta,\phi) \sim G(\theta,\phi)$. The density function is:

$$p(Y/\psi,\phi) = (\theta\phi)^{\phi} x^{\phi-1} e^{-\phi\theta x}$$

$$= \exp \phi\{lg\theta-\theta y\} \quad b(y,\phi)$$

where: $b(x,\theta) = \phi^{\phi} y^{\phi-1}/\Gamma(\phi)$; ϕ is a index.

$a(\theta) = lg(\theta)$, which implies: $a'(\theta) = 1/\theta$ and

$a''(\theta) = 1/\theta^2$, so that:

$E(Y/\theta,\phi) = \theta^{-1}$ and $v(Y/\theta,\phi) = 1/\theta^2$.

(c) Poisson Distribution.

The density in this case is:

$$p(Y/\mu) = \frac{\mu^y e^{-y}}{y!}, \qquad y=0,1,2...; \qquad \mu>0$$

$$= \exp[y \ lg \ \mu-y] \frac{1}{y!}, \qquad \text{where:}$$

$a(\theta) = e^{\theta}$, $\theta = lg\ \mu$ and $\phi=1$. Then

$E[Y/\theta,\phi] = a'(\theta) = e^{\theta} = \mu$ and $v[Y/\theta,\phi] = \frac{a''(\theta)}{\phi} = \mu$

(d) Binomial Distribution.

Let $(Y/p,n) \sim B(n,p)$. The density function in this case is:

$$p[Y/p,n) = \binom{n}{y} p^y(1-p)^{n-y}; \quad 0<p<1, \quad y=0,1,\ldots,n$$

$$= \mathrm{Exp}\left[y \lg\left\{\frac{\mu}{n-\mu}\right\} + n \lg(n-\mu)\right] \quad b(y,\phi)$$

with $np=\mu$, and:

$$\psi = \lg \frac{\mu}{1-\mu} \; ; \; a(\psi) = - n \lg n + n \lg(1+e^{\psi}).$$

Then $a'(\psi) = n \dfrac{e^{\psi}}{1+e^{\psi}}$ and $a''(\psi) = n \dfrac{e^{\psi}}{(1+e^{\psi})^2}$, and so:

$$E(Y/\psi,\phi) = \mu = np \text{ and } v(Y/\psi,\phi) = np(1-p)$$

5.2.5: Comments

At this point it is worth making the following comments:

(i) Wedderburn (1976) gives some properties of the MLE for the four standard members of the exponential family of error distribution handled by GLM combined with a variety of link functions. In particular, the estimates are unique for link functions giving sufficient statistics, for example the log-linear model in the Poisson case and the Logit model in the Binomial case.

(ii) Although the GLM framework provides a class of models suitable for application in many different problems it provides only approximate inferences based on asymptotic results mainly due to the complexity of the resulting likelihood.

(iii) When the data are continuous (for eg. normal and gamma) learning about ϕ provides information about variation in the data that is not accounted for by the regression of the $\underline{x}$'s on $\underline{\lambda}$'s. The adequacy of particular member of the general exponential family can be examined by making inference about ϕ [West, (1983)].

By contrast the discrete distribution do not have this flexibility for modelling the random variation; the binomial and Poisson models have ϕ fixed. The scale parameter can be used to model extra variation in the data. West (1983) reviews some approaches to this problem as, for example, the beta-binomial models of Crowder (1978) in the context of biological bio-essay and the approximate analysis developed by Williams (1982), and extend these analyses considering the general exponential family with the scale parameter ϕ unrestricted ($\phi>0$). The scale factor can be interpreted as simple discount factor to the likelihood when $\phi<1$, and as an expansion factor when $\phi>1$.

(iv) Since our aim is to extend the GLM to include dynamic models in which, of course, asymptotic results do not apply, it will be necessary to look for some other way of providing precise inferences about regression parameters and future observations [West, Harrison and Migon (1983)]. In Section 5.3 we discuss the extention of GLM to the dynamic linear case.

5.3: Dynamic Generalized Linear Model.

5.3.1: Introduction.

In order to extend the generalized linear model it is natural to replace $\underline{\theta}$ throughout by $\underline{\theta}_t$ and to adopt a linear time evolution for $\underline{\theta}_t$ as in DLM. A class of models with these features is introduced in this section.

A basic condition to get a tractable model is to use the closed Bayesian analysis. The conjugate family of distribution is discussed and the forecast distribution in standard form is presented.

It is worth pointing out once more that the main advantage of this class of models is in its use of the appropriate non-normal sampling distribution. Sampling models in the one parameter exponential family of distributions for steady models, that is models in which $\underline{\theta}_t$ is one dimensional with $\underline{\theta}_t = \mu_t$ for all t, are discussed in Smith [1979] and related approaches are taken by Souza [1981] and Azzalini [1983].

Although the analysis for the dynamic regression model is somewhat more complex, a tractable solution is presented in this section, and the same method of analysis may be used for a much wider class of non-linear models [Migon and Harrison (1983)] to be introduced in section 5.4.

5.3.2: The model structure.

The general exponential family sampling model for the observation Y_t is specified by the density:

$$p(Y_t/\psi_t,\phi) = \exp[\phi\{Y_t\psi_t - a(\psi_t)\}]\ b(Y_t,\phi) \qquad (5.3.1)$$

with ψ_t the natural parameter of the distribution, which satisfies:

$$E[Y_t/\psi_t,\phi] = a'(\psi_t) = \mu_t \text{ and } v(Y_t/\psi_t,\phi) = a''(\psi_t)/\phi \qquad (5.3.2)$$

where ϕ is a scale factor.

The prior distribution for the natural parameter has the conjugate form

$$p(\psi_t/D_{t-1},\phi) = c(\alpha_t\,\beta_t)\ \exp[\alpha_t\psi_t - \beta_t a(\psi_t)] \qquad (5.3.3)$$

and it will be denoted as $(\psi_t/D_{t-1},\phi) \sim CP[\alpha_t,\beta_t]$. Using the usual conjugate analysis we obtain the posterior distribution as:

$$(\ _t/D_t,\phi) \sim CP[\alpha_t + \phi\ y_t;\ \beta_t + \phi] \qquad (5.3.4)$$

and the predictive distribution is:

$$\begin{aligned} p(y_t/D_{t-1}) &= \int p(y_t/\psi_t)\ p(\psi_t/\phi,D_{t-1})\ d\ \psi_t \\ &= b(y_t,\phi)\ C(\alpha_t,\beta_t) \int \exp[\{\phi y_t+\alpha_t\}\ \psi_t - (\beta_t+\phi)a(\psi_t)]\ d\ \eta_t \\ &= b(y_t,\phi)\ C(\alpha_t,\beta_t)\ \frac{1}{C\big((\alpha_t+\phi y_t);\ \beta_t+\phi)\big)} \end{aligned}$$

The natural extension of the static generalized linear model consists of defining the systematic component of the model as a DLM, that is:

$$g(\psi_t) = \lambda_t = \underline{F}_t \, \underline{\theta}_t \tag{5.3.6}$$

Suppose furthermore that $(\underline{\theta}_{t-1}/D_{t-1}) \sim [\underline{m}_{t-1}, \underline{C}_{t-1}]$ and, by analogy with the DLM, define the time evolution of the model on the first two moments of the state variables as:

$$(\underline{\theta}_t/D_{t-1}) \sim [\hat{\underline{\theta}}_t, \underline{R}_t], \text{ where:}$$

$$\hat{\underline{\theta}}_t = \underline{G} \, \underline{m}_{t-1}$$

$$\underline{R}_t = \underline{B}_t^{-\frac{1}{2}} \, \underline{G}_t \, \underline{C}_{t-1} \, \underline{G}'_t \, \underline{B}_t^{-\frac{1}{2}} \tag{5.3.6}$$

with the transition matrix $\underline{G}_t$ and the discount matrix $\underline{B}_t$ known.

Note that the full distribution of the state vector is unspecified since the evolution is defined on the moments of the state vector. The natural restriction we impose is that the marginal distribution of λ_t, a linear function of the elements of $\underline{\theta}_t$, is such that ψ_t has the conjugate prior distribution of (5.3.3),

$$(\psi_t/D_{t-1}) \sim CP[\alpha_t, \beta_t]$$

From the linear structure of the model it follows that:

$$E[\lambda_t/D_{t-1}] = \underline{F}_t \, \hat{\underline{\theta}}_t = \hat{\lambda}_t$$

$V[\lambda_t/D_{t-1}] = \underline{F}_t \, \underline{R}_t \, \underline{F}'_t = r$, and the values of α_t and β_t are uniquely determined as function of $\hat{\lambda}_t$ and r. Finally, the covariance between λ_t and $\underline{\theta}_t$ is easily calculated as:

$$\underline{r}_t = \text{cov}[\lambda_t, \underline{\theta}_t/D_{t-1}] = \underline{R}_t \, \underline{F}_t$$

The posterior for (ψ_t/D_t) and the predictive distribution for (Y_t/D_{t-1}) are given by 5.3.4 and 5.3.5 respectively. In order to complete the full Bayesian analysis we would require the posterior distribution for $[\underline{\theta}_t/D_t]$, but the model, as developed so far, does not require the full specification of the prior for $(\underline{\theta}_t/D_{t-1})$, and so the posterior distribution for $(\underline{\theta}_t/D_t)$ is not available. The central point is that only the mean vector and the covariance matrix are required, and they satisfy the identities:

$$\underline{m}_t = E\ E\{\underline{\theta}_t/\psi_t, D_t\}$$

$$\underline{C}_t = V[\ E\{\underline{\theta}_t/\psi_t, D_t\}] + E[v\{\underline{\theta}_t/\psi_t, D_t] \qquad (5.3.7)$$

Since, as in the Normal case, $(\underline{\theta}_t/\psi_t, D_t)$ is conditionally independent of $(Y_t, \underline{F}_t)$, Chapter 3 section 3, all the information available to update the moments of $\underline{\theta}_t$, which are unknown non-linear function of ψ_t, is that concerning the joint moments of $(\lambda_t, \underline{\theta}_t/D_{t-1})$,

$$\left(\begin{array}{c|c} \lambda_t & \\ \underline{\theta}_t & D_{t-1} \end{array}\right) \sim \left[\begin{pmatrix} \hat{\lambda}_t \\ \hat{\underline{\theta}}_t \end{pmatrix}, \begin{vmatrix} r & \underline{r}_t \\ \underline{r}_t' & \underline{R}_t \end{vmatrix}\right]$$

In the coming section an approach to filter back the information $\{Y_t, \underline{F}_t\}$ is developed.

5.3.3 The posterior distribution for the state vector.

As we have discussed in Chapter 3 the main assumptions of our model are:

(i) $p[\underline{\theta}_t, \psi_t/D_t] = p[\psi_t/y_t, D_{t-1}]\ p[\underline{\theta}_t/\psi_t, D_{t-1}]$, i.e.

$\underline{\theta}_t$ is conditionally independent of Y_t;

(ii) the prior distribution for $\underline{\theta}_t$ given D_{t-1} is only partially specified; and

(iii) the marginal distribution for $\lambda_t = \underline{F}_t\ \underline{\theta}_t$ is such that the mean and variance of λ_t are linked to the parameters of the conjugate prior for $(\psi_t/D_{t-1}) \sim CP(\alpha_t, \beta_t)$

From assumption (ii) it is clear that our model does not satisfy the requirement for a full Bayesian analysis, that is we cannot get the posterior for $(\underline{\theta}_t/D_t)$. On the other hand, all we need is to calculate the mean and covariance of this random vector.

Since the mean of $(\underline{\theta}_t/\psi_t, D_{t-1})$ is the optimal predictor in the sense of minimizing the quadratic risk function, and the covariance matrix is just the value of $\underline{A}(\underline{d}) = E[(\underline{\theta}-\underline{d})(\underline{\theta}-\underline{d})'/\psi_t, D_{t-1}]$ at the maximum, our problem is, in resume, to find this mean value. Note that $p(\underline{\theta}_t/\psi_t, D_{t-1})$ is an unknown predictive distribution of $\underline{\theta}_t$ given ψ_t, and that the mean is not known. We must use some alternative approach to predict it.

The linear Bayes method was used by Hartigan (1969) to problems of inference in linear models when only the first two moments of the prior and likelihood are specified, rather than the complete probability model. A similar approach is discussed by Goldstein (1976) in the context of non-linear regression problem.

The application the the linear Bayes approach to our case is straightforward and it provides the feedback of the information in $(Y_t, \underline{F}_t)$ to $\underline{\theta}_t$.

Let $\underline{d}$ be a linear predictor defined as: $\underline{d} = \underline{d}_0 + \underline{d}_1 \lambda$ and let the overall quadratic risk be given by:

$$r_t(\underline{d}) = \text{trace } E[\underline{A}_t(\underline{d})/D_{t-1}] \tag{5.3.8}$$

where: $\underline{A}_t(\underline{d}) = E[(\underline{\theta}_t-\underline{d})(\underline{\theta}_t-\underline{d})'/\psi_t, D_{t-1}]$.

The minimization of 5.3.8 gives:

$$\begin{cases} \hat{\hat{\underline{m}}}_t = \underline{m}_t + \underline{r}_t' [\lambda_t - \hat{\lambda}_t]/r_t \\ \hat{\underline{R}}_t = E[\underline{A}_t(\underline{d})/D_{t-1}]\Big|_{\underline{d}=\hat{\hat{\underline{m}}}_t} \end{cases}$$

$$= \underline{R}_t - \underline{r}_t'\underline{r}_t/r_t \tag{5.3.9}$$

It is worth noting that 5.3.9 are exactly the mean vector and the covariance matrix of $(\underline{\theta}_t/\psi_t, D_{t-1})$ in the normal model (Chapter 3) and so will be close to the true moments in non-normal models if $(\psi_t, \theta_t/D_{t-1})$ is approximately normal. To complete the feedback of information we need to substitute the conditional mean and variance matrices in (5.3.7) by $\hat{\hat{\underline{m}}}_t$ and $\hat{\underline{R}}_t$ to obtain:

$$\underline{m}_t = \hat{\underline{m}}_t + \underline{r}_t'[g_t - \hat{\lambda}_t]/$$

$$\underline{C}_t = \underline{R}_t - \underline{r}_t'\underline{r}_t[r_t - p_t]/r_t^2, \quad \text{where}$$

$g_t = E[\lambda_t/D_t]$ and $p_t = v[\lambda_t/D_t]$ are calculated from the fully specified conjugate posterior for (ψ_t/D_t) using the appropriate link function.

Now we have the complete updating system for the dynamic Generalized Linear Model and in the coming paragraph we present some operational aspects of the proposed method for special cases. Numerical examples are presented in West, Harrison and Migon (1983).

5.3.4 Special cases and applications.

Two special cases of importance are the discrete binomial and Poisson Models. As we have pointed out before, suitable models are often obtained taking $g(\cdot)$ to be the identity function, so that the linear evolution is defined for the natural parameter, i.e: $\psi_t = \lambda_t$. This leads to the logistic-linear and log-linear models in the binomial and Poisson cases respectively.

(a) Poisson Model:

For this member of the exponential family $\psi_t = \log(\mu_t)$, where μ_t is the mean or Poisson rate. Since $\mu_t > 0$, then ψ_t is a real value.

The conjugate prior for $(\psi_t/D_{t-1}) \sim CP[\alpha_t, \beta_t]$ is such that (μ_t/D_{t-1}) has a gamma distribution and the parameters (α_t, β_t) are functions of $(\hat{\lambda}_t, r_t)$, with

$$\hat{\lambda}_t = E[\lg(\mu_t)/D_{t-1}] = \int_0^\infty \lg \mu_t \, \mu_t^{\alpha_t - 1} \, e^{\beta_t \mu_t} \frac{\beta_t^{\alpha_t}}{\Gamma(\alpha_t)} \, d\mu_t$$

$$= \gamma(\alpha_t) - \lg(\beta_t)$$

$$r_t = v[\lg(\mu_t/D_{t-1})] = \gamma'(\alpha_t) \qquad (5.3.10)$$

where: $\gamma(\alpha) = \frac{d}{d\alpha} \lg \Gamma(\alpha)$, $\Gamma(\cdot)$ - gamma function, is the digamma function.

Therefore, of course, some approximation and recurrence relations in order to get α_t and β_t as function of $(\hat{\lambda}_t, r_t)$. From Abramovitz, M. and Stegun, I.A. handbook we have;

(i) recurrence relationships:

$$\begin{cases} \gamma(\alpha) = \gamma(\alpha+1) - \frac{1}{\alpha} \\ \gamma'(\alpha) = \gamma'(\alpha+1) + \frac{1}{\alpha^2} \end{cases}$$

(ii) approximation formulas:

$$\begin{cases} \gamma(\alpha) \simeq \lg(\alpha) - \frac{1}{2\alpha} \\ \gamma'(\alpha) \equiv \frac{1}{\alpha} + \frac{1}{2\alpha^2}, \end{cases}$$ which are based on Stirling formulas.

A better approximation for the digamma function is:

$$\gamma(\alpha) = \lg(\alpha - .5) \quad \text{for } \alpha > 3.0.$$

Using this approximation with $r_t < .4$ the values of α_t and β_t follow:

$$\begin{cases} \alpha_t = r_t^{-1} + .5 \\ \beta_t = r_t^{-1} + \exp(-\hat{\lambda}_t) \end{cases}$$

Finally the predictive distribution is:

$$p(Y_t/D_{t-1}) = \int_0^\infty p(Y_t/\psi_t) \, p(\psi_t/D_{t-1}) \, d\,\psi_t$$

$$= \begin{pmatrix} \alpha_t + y_t - 1 \\ y_t \end{pmatrix} (1-\ell)^{y_t} \, \ell^{\alpha_t},$$ where

$\ell = \frac{\beta_t}{1+\beta_t}$; $y_t = 0,1,2,\ldots$; that is a negative-binomial. The forecast for Y_{t+k}, given data up to time t, is taken to be the mode of the one step ahead negative binomial predictive distribution.

(b) Binomial model.

In this example the natural parameter ψ_t is linked to the parameter μ_t of the binomial as:

$$\psi_t = \lg\left[\frac{\mu_t}{1-\mu_t}\right] \text{; which is the logistic transformation.}$$

The parameter μ_t has a beta prior distribution with:

$$\begin{cases} \hat{\lambda}_t = \gamma(\alpha_t) - \gamma(\beta_t) \\ r_t = \gamma'(\alpha_t) + \gamma'(\beta_t), \end{cases}$$ and using the approximations described above we get:

$$\begin{cases} \alpha_t = \frac{e^{\hat{\lambda}}}{r_t} + .5 \\ \beta_t = r_t^{-1} + .5 \end{cases}$$

The predictive distribution is a Beta-Binomial distribution and we suggest the use of the mean value as a lead time point forecast.

5.4 A Class of Dynamic Non-Linear Models.

In this section we extend further the model described in 5.3. A useful class of Dynamic non-linear models is introduced as a generalization of the Normal dynamic non-linear models of Chapter 3 [Migon and Harrison (1983)].

This class of models relate the mean of the observational model with the state vector through a general link function, often non-linear, as was done in Chapter 3 for the normal case. We do not need to be restricted to the exponential family of distribution, neither to the conjugacy analysis, even though the analysis is much simpler in this case.

This class of models offers a wider extention of the dynamic generalized linear model. The breaking back of information is justified by two different ways and some examples are presented.

5.4.2 The guide relationship.

Let $\underline{\theta}_t$ be a parameter vector with $[\underline{\theta}_{t-1}/D_{t-1}] \sim [\underline{m}_{t-1}, \underline{C}_{t-1}]$. Suppose that a specific dynamic model, which may be non-linear, applied to this posterior distribution gives the prior $(\underline{\theta}_t/D_{t-1}) \sim [\hat{\underline{\theta}}_t, \underline{R}_t]$. Note that posterior and prior for $\underline{\theta}_t$ are partially specified.

Let ψ_t be a parameter related to $\underline{\theta}_t$ through the guide relationship $g(\cdot)$, which can be deterministic or stochastic, such that:

$$\left(\begin{matrix}\psi_t \\ \underset{\sim}{\theta}_t\end{matrix}\middle| D_{t-1}\right) \sim \left[\begin{pmatrix}\hat{\psi}_t \\ \hat{\theta}_t\end{pmatrix} ; \begin{bmatrix} r_t & \underline{r}_t \\ \underline{r}'_t & \underline{R}_t\end{bmatrix}\right] \sim [\hat{\phi}, \underline{U}_t]$$

and,

$$\psi_t = g(\underline{\theta}_t) + \delta_t,\ \delta_t \sim [0;\sigma^2]$$

where δ_t introduces some extra variation. Again, the joint prior for $(\psi_t, \underline{\theta}_t)'$ is only partially specified.

The marginal distribution of ψ_t is defined for each application, say, as $p(\psi_t/D_{t-1})$. It will often be taken to have a conjugate form to that of the observation Y_t or to be such that a bijective function of ψ_t has the conjugate distribution.

As in the dynamic GLM, we assume that the information that Y_t gives on $\underline{\theta}_t$ is conveyed only through ψ_t, and the posterior of ψ_t is derived by Bayes theorem as:

$$\lg p(\psi_t/y_t, D_{t-1}) = S(\psi_t/y_t) + \lg p(\psi_t/D_{t-1}) + \text{const.}$$

where: $S(\psi_t/y_t)$ is the log-likelihood of (ψ_t/y_t).

Let the posterior mean and variance of $(\psi_t/y_t,D_{t-1})$ be

$$E(\psi_t/D_t) = m \text{ and } v(\psi_t/D_t) = \sigma_{11}$$

In the coming section we discuss the breaking back of the information.

5.4.3 Updating recurrence for the state parameters.

The linear Bayes procedure of Section 5.3.4 can be used straightforwardly. Note that the structure of both models is the same; so the updating relationships are:

$$\begin{cases} \hat{\underline{m}}_t = \hat{\underline{\theta}}_t + \underline{r}_t\,[\psi_t-\hat{\psi}_t]/r_t \\ \hat{\underline{R}}_t = R_t - \underline{r}_t'\,\underline{r}_t/r_t \end{cases} \qquad \text{where } \hat{m}_t = E[\underline{\theta}_t/D_{t-1},\psi_t] \text{ and } \hat{\underline{R}}_t = V(\underline{\theta}_t/D_{t-1},\psi_t).$$

Taking the expectation, as in 5.3.7, we get:

$$\begin{cases} \underline{m}_t = \hat{\underline{\theta}}_t + \underline{r}_t'\,[m-\hat{\psi}_t\,/r_t] \\ \underline{C}_t = \underline{R}_t - \underline{r}_t'\,\underline{r}_t\,[r_t-\sigma_{11}]/r_t^2 \end{cases} \tag{5.4.2}$$

The identities in 5.4.2 are exactly the recurrence obtained in the normal case. The linear Bayes solution, apart from providing the optimal predictor in the sense of minimizing the quadratic risk function, is the exact solution for the normal case. In any case we have a reasonable approximation for the first moments of the posterior distribution of $\underline{\theta}_t$.

A different justification for the identities in 5.4.2 is based on the following optimization principles:

(i) Let $\underline{m}$ be the value of $\underline{\alpha}$ which minimizes

$$L(\alpha) = (\hat{\underline{\phi}}-\underline{\alpha})'\,\underline{U}_t^{-1}\,(\hat{\underline{\phi}}-\underline{\alpha}) \;,\; \underline{U}_t = \begin{bmatrix} r_t & \underline{r}_t \\ \underline{r}_t' & \underline{R}_t \end{bmatrix}$$

subject to the constraint $\alpha_1=m$, the first component of $\underline{\alpha}$; and

(ii) Let $\underline{\Sigma} = (\sigma_{ij})$ be that variance matrix $\underline{M}$ which minimizes

$$L(\underline{\Sigma}) = |\underline{\ell}'\{\underline{M}^{-1} - (\underline{U}_t^{-1} + \begin{bmatrix} \sigma_{11}^{-1} & -r^{-1} & \underline{0} \\ & \underline{0} & \underline{0} \end{bmatrix})\}\,\underline{\ell}|$$

$\forall\ \underline{\ell} \in \mathbb{R}^n$, subject to the leading element being σ_{11}.

Some interesting comments are as following:

(i) Note that $\frac{1}{\sigma_{11}} - \frac{1}{r}$ is the gain in precision for ψ_t from prior to posterior

(ii) It is easily checked that, $\underline{C}_t = R_t^{-1} + \begin{bmatrix} c_{11}^{-1} & -r_{11}^{-1} & \underline{0} \\ \underline{0} & & \underline{0} \end{bmatrix}$ is such that $L(\underline{C}) = 0$;

(iii) From (i) follows, using the Lagrangian multipliers, that

$$\frac{\partial}{\partial \underline{\alpha}} L(\underline{\alpha}) = Z\, \underline{U}^{-1}\,(\hat{\phi}-\underline{\alpha}) + \lambda \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0$$

$$\frac{\partial}{\partial \lambda}\,(\underline{\alpha}) = \alpha_1 - m = 0$$

so:

$$\underline{U}^{-1}\,\underline{\alpha} = \underline{U}^{-1}\,\hat{\phi} + \frac{\lambda}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{or}$$

$$\underline{\alpha} = \begin{pmatrix} \alpha_1 \\ \underline{\alpha}_2 \end{pmatrix} = \begin{pmatrix} \hat{\psi} \\ \hat{\underline{\theta}} \end{pmatrix} + \frac{\lambda}{2} \begin{pmatrix} r \\ \underline{r} \end{pmatrix}$$

Then using the constraint equation we got:

$\frac{\lambda}{2} = (m-\overline{\psi})/r$, and, finally:

$$\underline{\alpha} = \begin{bmatrix} \hat{\psi} \\ \hat{\underline{\theta}} \end{bmatrix} + \frac{m-\overline{\psi}}{r} \begin{bmatrix} r \\ \underline{r} \end{bmatrix} .$$

5.4.4 Examples

(a) The log-normal model.

Let Y_t be distributed as a log-normal, $Y_t \sim LN(\mu,\sigma)$, where μ and σ are the mean and variance of the associated normal distribution. It is well known but worth remembering that:

$$M = E(Y) = \exp\{\mu + \frac{\sigma}{2}\}$$

$S = v(Y) = M^2\{\exp(\sigma)-1\}$; and the square of the coefficient of variation is:

$$K^2 = \operatorname{Exp}(\sigma)-1$$

The dynamic non-linear model may be easily applied to this probabilistic situation as follows:

(i) Observational model and joint prior

- $(Y_t/\psi_t) \sim LN[\mu_0,\sigma_0]$, with $\psi_t = E(Y_t/\psi_t,\sigma_0)$

- $\left[\begin{matrix}\psi_t \\ \underline{\theta}_t\end{matrix}\middle| D_{t-1}\right] \sim \left[\begin{pmatrix}\hat{\psi}_t \\ \hat{\underline{\theta}}_t\end{pmatrix}; \begin{bmatrix} r & \underline{r} \\ \underline{r}' & \underline{R}_t\end{bmatrix}\right]$, where:

as usual $\hat{\psi}_t = [E\ g(\underline{\theta}_t/D_{t-1})]$; $r = \operatorname{var}[g(\underline{\theta}_t)/D_{t-1}]$ and $\underline{r} = \operatorname{cov}[g(\underline{\theta}_t), \underline{\theta}_t/D_{t-1}]$.

(ii) Let the marginal distribution $(\psi_t/D_{t-1}) \sim LN[m_0,s_0]$, where:

$$\begin{cases} m_0 = \lg(\hat{\psi}_t) - \frac{1}{2}\lg\left(\frac{r-\hat{\psi}_t^2}{\hat{\psi}_t^2}\right) \\ s_0 = \left\{\lg\left[\frac{r-\hat{\psi}_t^2}{\hat{\psi}_t^2}\right]\right\}^2 \end{cases}$$

(iii) The posterior marginal distribution of ψ_t given D_t is

$(\psi_t/D_{t-1},y_t) \sim LN[m_1,\sigma_1]$, with

$$\begin{cases} m_1 = m_0 + [\lg y_t - m_0]\frac{s_0}{s_0+\sigma_0} \\ s_1 = s_0 - \frac{s_0^2}{s_0+\sigma_0} \end{cases}$$

and it follows that:

$$\begin{cases} E(\psi_t/D_t) = m = \exp\left(m_1 + \frac{s_1}{2}\right) \\ v(\psi_t/D_t) = \sigma_{11} = \{\exp(s_1)-1\}\, m^2 \end{cases}$$

(iv) The updating for the underlying parameter is

$$(\underline{\theta}_t/D_t) \sim [\underline{m}_t, \underline{C}_t] \text{ , where:}$$

$$\begin{cases} \underline{m}_t = \underline{\hat{\theta}}_t + \underline{r}'[m-\hat{\psi}_t]/r \\ \underline{C}_t = \underline{R}_t - \underline{r}'\ \underline{r}\,[r-\sigma_{11}]/r^2 \end{cases}$$

If the variance of (Y_t/ψ_t) is unknown we can learn about it on-line using, as in Chapter 3,

$$d_t^2 = (1-A_t)\ e_t^2 \ , \qquad A_t = \frac{s_0}{s_0+\sigma_0}$$

$$\begin{cases} \alpha_t = \alpha_{t-1} + d_t^2 \\ n_t = n_{t-1} + 1 \end{cases} \qquad \text{and defining}$$

$$\hat{\sigma}_0^2 = \alpha_t/n_t$$

It is worth pointing out that a discount factor, β_v, can be used for the variance learning procedure and that some protection against outliers can be introduced as in Chapter 3.

a.1) A numerical example.

We present a seasonal growth model with log-normal observational distribution and some numerical results for two data sets.

Let the evolution of $\underline{\theta}_t$ be:

$$(\underline{\theta}_t/D_{t-1}) \sim [\underline{\hat{\theta}}_t;\ \underline{B}^{-\frac{1}{2}}\ \underline{G}\ \underline{C}_{t-1}\ \underline{G}'\ \underline{B}^{-\frac{1}{2}}]$$

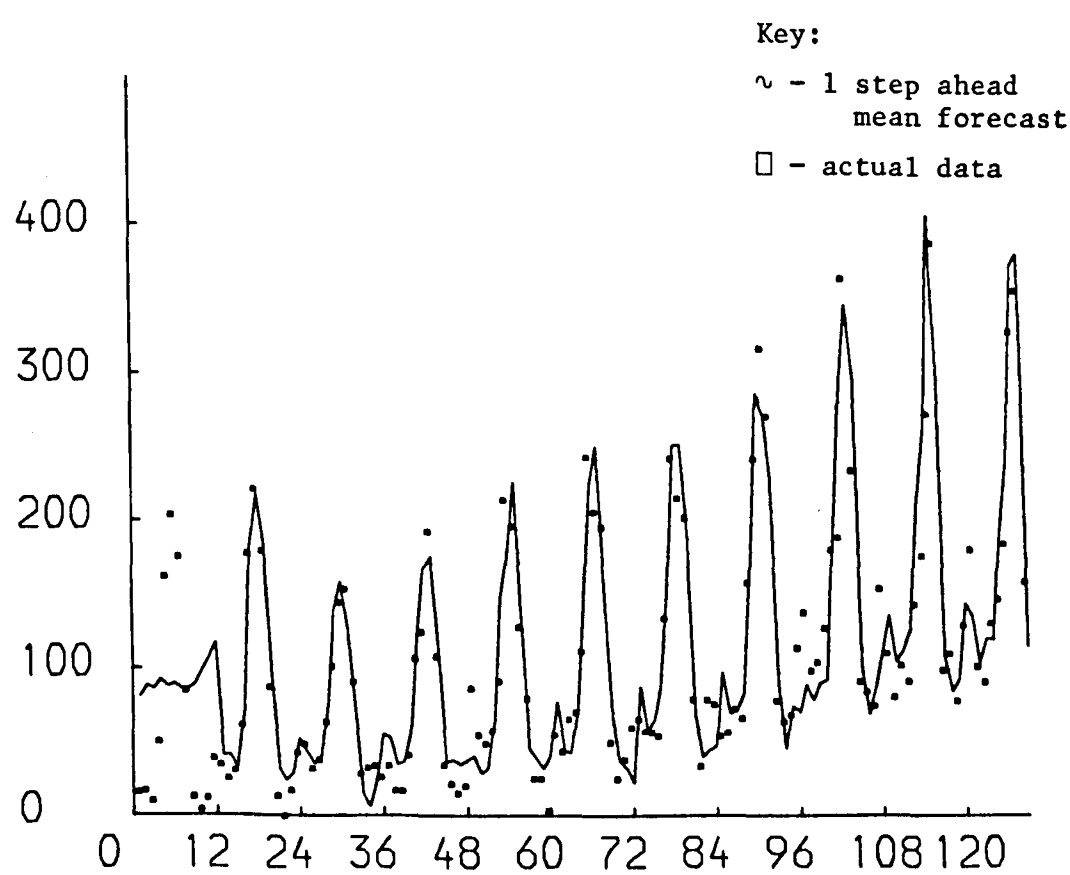


Fig. 1: Turkey poult sales (*) and one-step-ahead mean forecast. β_1 = .95; β_2 = .93.

where:

$$\underline{G} = \mathrm{diag}(\underline{G}_1\ \underline{G}_2 \ldots G_7),\ \underline{G}_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

$$\underline{G}_{i+1} = \begin{bmatrix} \cos(iw) & \sin(iw) \\ -\sin(iw) & \cos(iw) \end{bmatrix},\ w = \frac{2\pi}{12}, \quad i=1,\ldots 6, \text{ and}$$

$$\underline{B} = \mathrm{diag}(\beta_1\ I_2,\ \beta_2\ I_{12\times 12})$$

Now consider the guide relationships:

$$\psi_t = (\underline{F}\ \underline{\theta}_t) * (1+\underline{H}\ \underline{\theta}_t), \quad \text{with}$$

$$\underline{F} = (1,0\ldots 0) \quad \text{and} \quad \underline{H} = (0,0,1,0,1,0,\ldots,1,0)$$

The 129 observations plotted in Fig. 1 are the month sales data of turkey poult, as in Chapter 3. The full line in Fig. 1 provides the one-step-ahead mean value. At observation number 29 we have intervened in the model dropping the level of the process by 20 units. As in Chapter 3 we have fitted a seasonal model with 5 harmonics, and we have used the same values for the discount factor, that is $\beta_1=.95$ and $\beta_2=.93$.

As a second example for the log-normal model we used the monthly sales data of an engineering company discussed by Chatfield and Prothero (1973) and West-Harrison and Migon (1983). The log-normal seasonal growth model with 6 harmonics performs very well. In Fig. 2 we present the one-step-ahead forecast for the period January 1965, as t=1, to May 1971 as t=77, and then forecasts for the next 12 months from t=77 using this model.

The year ahead forecast in Fig. 2 are lower than those of the models mentioned in Chatfield and Prothero. For t=78 to 84 the actual observations and forecasts from t=77 are given by:

time	78	79	80	81	82	83	84
observation	260	304	390	614	783	872	540
forecast	253	336	434	664	882	927	709

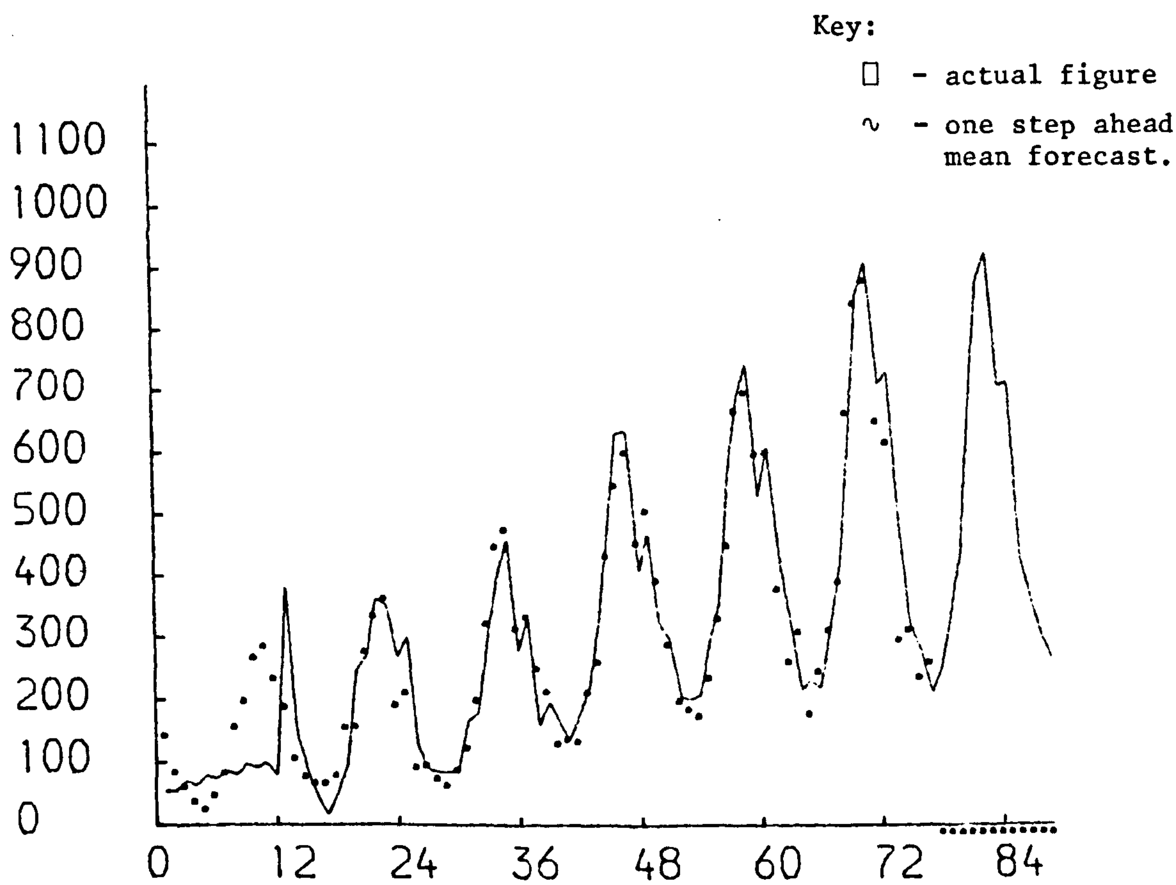


Fig. 2: Engineering Co. Sales data (*) and one-step-ahead mean forecast $\beta_1 = .85$; $\beta_2 = .97$

b) The Beta-Binomial case.

In this section we present an illustration of the dynamic non-linear model to the Beta-Binomial case, and in particular to the non-linear model - model 1 - of Chapter 4.

(i) Let $(\underline{\theta}_{t-1}/D_{t-1}) \sim [\underline{m}_{t-1}, \underline{C}_{t-1}]$. Then $(\underline{\theta}_t/D_{t-1}) \sim [\hat{\underline{\theta}}_t, \underline{R}_t]$ is obtained via the relationship, $\underline{\theta}_t = G\, \underline{\theta}_{t-1} + \underline{w}_t$, $\underline{w}_t \sim [0, \underline{W}]$;

(ii) Consider the following guide relationships:

$$\psi_t = p_t + \delta_t, \qquad \delta_t \sim [0, \sigma^2]$$

$$p_t = \mu_t + \lambda_t\, E_{t-1} + [d_t - \lambda_t\, E_{t-1}]\, k_t\, x_t,$$

where the notation is as in Chapter 4.

(iii) The joint prior distribution is obtained as:

$$\left(\begin{matrix}\psi_t \\ \underline{\theta}_t\end{matrix}\middle| D_{t-1}\right) \sim \left[\begin{pmatrix}\hat{\psi}_t \\ \hat{\theta}_t\end{pmatrix} ; \begin{bmatrix} r & \underline{r} \\ \underline{r}' & \underline{R}\end{bmatrix}\right], \quad \text{where:}$$

$r = v(\psi_t/D_{t-1})$, $\underline{r} = \mathrm{cov}(\psi_t, \underline{\theta}_t/D_{t-1})$, and

$\hat{\psi}_t = E(\psi_t/D_{t-1})$, are calculated using the formulas for the general moments of a normal distribution.

(iv) Let the marginal distribution $(\psi_t/D_{t-1}) \sim Be(s,t)$, where:

$$\begin{cases} s = \hat{\psi}_t[\hat{\psi}_t(1-\hat{\psi}_t) - r]/r \\ t = s(1-\hat{\psi}_t)/\hat{\psi}_t \end{cases}$$

ensures that the first two moments are $(\hat{\psi}, r)$ and

$p(\psi_t = \psi/D_{t-1}) \propto \psi^{s-1}\ (1-\psi)^{t-1}$.

(v) Let $(Y_t/\psi_t, N_t) \sim B(N_t, \psi_t)$ so that the likelihood of ψ_t given Y_t is:

$$L(\psi_t/y_t,N_t) \propto \psi_t^{y_t}\ (1-\psi_t)^{N_t-y_t}$$

(vi) Using Bayes' theorem the posterior for ψ_t follows:

$$(\psi_t/D_t) \sim Be(s_1,t_1), \quad \text{where:}$$

$$\begin{cases} s_1 = s+y_t \\ t_1 = t+N_t-y_t \end{cases}$$

So that the mean and variance are:

$$m = \frac{s_1}{s_1+t_1}$$

$$\sigma_{11} = m(1-m)/(s_1+t_1-1)$$

(vii) The updating equation for the state vector, $\underline{\theta}_t$, is:

$$(\underline{\theta}_t/D_t) \sim [\underline{m}_t,\underline{c}_t]$$

$$\begin{cases} \underline{m}_t = \hat{\underline{\theta}}_t + \underline{r}'\ [m-\hat{\psi}_t]/r \\ \underline{C}_t = \underline{R}_t - \underline{r}'r\ \ [r-\sigma_{11}]/r^2 \end{cases}$$

It is worth pointing out that the random variable δ_t introduces any required extra variation arising from the sample TV region being unrepresentative of the country as a whole.

Finally in Fig. 3 the data for product P3 is given. Both the observed weekly number of people aware in a sample of 66 and the weekly T.V.R. are plotted for a period of more than three years. Periods of missing observations are indicated and the one week ahead point forecast, in this example the prior mean, are plotted. In this example N_t was assumed fixed and δ_t identically zero.

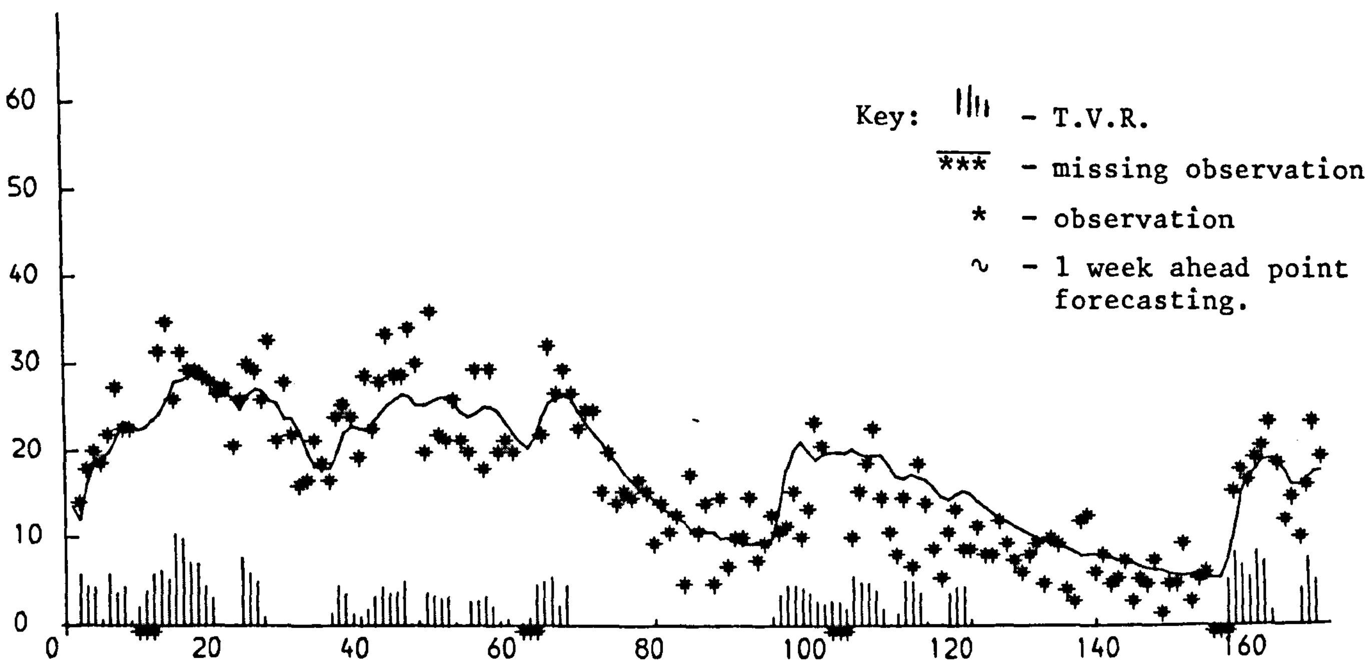


Fig. 3: Weekly number of people aware (*) - Product P3 - and one week ahead point forecast.

Chapter 6 Transfer Response; modelling and estimation.

6.1 Introduction

This chapter is concerned with the on-line extimation and modelling of transfer response. Over the last two decades a great effort has been concentrated on this problem by statisticians, system theorists and econometricians.

In the chapters 11 and 12 of Box-Jenkins (1970) the bases of transfer functions models are stated. It is worth pointing out that the unknown parameters are assumed constant in their analysis. Some very important applications to analysis of intervention can be found in Box-Taio (1975), and a good survey of the methods that have been developed primarily in system theory is found in Astrom and Eykhoff (1971). In econometrics the rational lag models are discussed in the book of Harvey (1981) and the works of Sarris (1973) and Cooley and Prescott (1976) are related to our work.

Our aim is to show how a transfer response can be modelled as a full probability distribution and to apply the results presented in Chapters 3 and 5 to the on-line estimation of stochastic transfer response.

The DLM structure is used and the principle of superposition for linear models permits the addition of an "effect model" to a base model. There is, of course, a great deal of scope for mathematically modelling any subjective opinion on form effects and additional uncertainty. For example in Chapter 4 the base model is a first order polynomial (the level) and the effect model was developed based on the hypothesis of diminishing returns from advertising and an exponential decay for awareness.

In the next section we present the classical models and discuss the Box-Jenkins approach. The Bayesian method is described in section 6.3 and some simulation results presented in section 6.4.

6.2 Transfer function modelling and estimation.

An important problem in many areas of application is that of the identification or estimation of a transfer response. For example in economics, engineering or physical sciences we frequently meet this sort of problem. Much of the literature assumes that the system can be adequately approximated over the range of interest by a linear model whose parameters do not change with time. Our interest is with models with parameters varying through the time, and we consider only systems in open-loop form in order to avoid discussions about the direction of causality.

6.2.1 Classical formulation

Suppose that the variations of Y_t and X_t over the range of interest can be well approximated by a linear steady-state relationship:

$$y_t = g.x_t \tag{6.2.1}$$

where g is the gain factor, and, y_t and x_t represent the variations of Y_t and X_t.

In general with x_t varying along the time we can represent the system by a linear filter of the form.

$$y_t = v_0\, x_t + v_1\, x_{t-1} + v_2\, x_{t-2} + \ldots$$

or

$$y_t = v(B)\, x_t \tag{6.2.2}$$

where B is the back shift operator, i.e.: $B\, y_t = y_{t-1}$.

The weights $\{v_0\ v_1\ v_2 \ldots\}$ are called the impulse response function of the system. These weights provide a description of the system in the time domain. For example, suppose that the input x_t is zero for all t except at time t_0 when it takes the value unity, i.e.: $x_{t_0} = 1$. Then the output at time t is given by $y_t = \sum_{k\geq 0} v_k\, x_{t-k} = v_{t-t_0}$.

An alternative way of describing a linear system in the time domain is by means of a function called the step response function. Suppose that $x_t = 0 \; \forall \; t<t_0$ and $x_t = 1 \; \forall \; t \geq t_0$ then:

$$s_t = \sum_{t_0 \leq k \leq t} v_{t-k} \, x_k = \sum_{t_0 \leq k \leq t} v_{t-k}$$

Since a discrete dynamic system is often parsimoniously represented by the general difference equation

$$\delta(B) y_t = \omega(B) x_t \tag{6.2.3}$$

where: $\delta(B) = 1 - \delta_1 B - \ldots \delta_r B^r$ and $\omega(B) = \omega_0 - \omega_1 B \ldots\ldots - \omega_s B^s$ are polynomials of degree r and s respectively. Assuming that $\delta(B)$ is invertible, it follows from (6.2.3) that the transfer function for this model is:

$$v(B) = \delta^{-1}(B) \, \omega(B)$$

This is called a rational transfer response. In general, for stability, we require that the roots of the characteristic equation $\delta(B) = 0$ with B regarded as a variable lie outside the unit circle.

The gain factor g is obtained from (6.2.3) holding x_t fixed, without loss of generality, at the value 1, as

$$g = \frac{\omega_0 - \omega_1 \ldots . - \omega_s}{1 - \delta_1 \ldots . - \delta_r}$$

Note that the time invariancy of the parameters implies that the gain factor is fixed. In the coming section we present the Bayesian formulation for transfer response and we show different ways of modelling the gain factor as a stochastic component.

6.2.2 The Box-Jenkins Modelling Procedure.

Generally, a Box-Jenkins transfer function analysis follows the steps of identification, estimation and diagnostic check as in the univariate case.

Identification:

(a) impose some particular transformation (say logarithmic or a Box-Cox transformation) upon the time series involved;

(b) fit an ARMA model to the (differenced) input. For example:

$$\delta(B)\ x_t = \omega(B)\ \alpha_t,\ \alpha_t \text{ purely random process.}$$

(c) transform the output using the same operators identified in (b)

$$\omega^{-1}(B)\ \delta(B)\ y_t = \beta_t,\ \beta_t \text{ purely random process;}$$

(d) calculate the cross-covariance function of the filtered input and output, namely $\{\alpha_t\}$, $\{\beta_t\}$:

$$\gamma_{\alpha\beta}(m) = v_m\ \mathrm{var}(\alpha_t)$$

(e) use the univariate identification procedure to establish a stochastic model for the effects unaccounted for random shocks, past as well as present.

Let $N_t = \frac{\theta(B)}{\phi(B)}\ (1-B)^d\ a_t$ where

$\theta(B)$, $\phi(B)$ are operator of degree q and p respectively; a_t is a normal random variable.

Estimation:

Maximum likelihood estimates are derived for the parameters in $\omega(B)$ and $\delta(B)$, for each model identified before, and in $\theta(B)$ and $\phi(B)$ for the corresponding noise model.

Diagnostic check:

The residuals, a_t, of the tentative model are checked to determine whether they are correlated with the unsystematic changes in the input. If the correlations are significant it indicates that the tentative transfer function is inadequately using the explanatory power of the input variable.

6.3 Bayesian Stochastic Transfer Response.

6.3.1 General.

As discussed in Chapter 2 the systematic dynamic component or trajectory of a DLM is completely characterized by the eigenvalues of the system matrix $\underline{G}$. A transfer function can be modelled in DLM form as

$$\text{observation equation} - Y_t = \underline{F}\ \underline{\theta}_t + \epsilon_t$$

$$\text{system equation} - \underline{\theta}_t = \underline{G}(\underline{\lambda})\ \underline{\theta}_{t-1} + \underline{\Gamma}\ x_t \qquad (6.3.1)$$

where: $\epsilon_t \sim IN(0,V_t)$ and, without loss of generality, we suppose that

$x_t = \begin{cases} x & \text{if} \quad t=1 \\ 0 & \text{if} \quad t\neq 1 \end{cases}$, and $(\underline{\theta}_0/D_0) \sim N(\underline{0},\underline{C}_0)$. With this setting the forecasting function is:

$$\begin{aligned} F_t(k) &= E[Y_{t+k}/D_t,x_t] \\ &= \underline{F}\ \underline{G}^k\ \underline{\Gamma}\ x,\ k\geq 0 \end{aligned} \qquad (6.3.2)$$

Note that only the case with ϵ_t independent normal is considered although any structure can be used for the residual component. If $\underline{\lambda}$ and $\underline{\Gamma}$ are known the DLM can be applied without problems. Our main interest in this chapter is to show how, using the results of Chapter 3, we can estimate sequentially the unknown value of $\underline{\lambda}$. First of all it is worth showing that the gain factor $\underline{\Gamma}$ can be modelled in many different ways.

6.3.2 Transfer Response Modelling.

In order to show different ways of modelling the gain factor we concentrate on a first order model.

(i) locally constant model.

Suppose that $\underline{\Gamma}_t = \underline{\Gamma}$; $\forall t$. This is just the classical model discussed in previous sections. For example the first order model with decay factor λ known and gain γ fixed and known can be written as:

$$\begin{cases} Y_t = \theta_t + \epsilon_t \\ \theta_t = \lambda\, \theta_{t-1} + \gamma\, x_t \end{cases}$$ or, equivalently as:

$$Y_t - \lambda\, Y_{t-1} = \gamma\, x_t + \epsilon_t - \lambda\, \epsilon_{t-1}$$

or

$$Y_t = \frac{\gamma}{1-\lambda B}\, x_t + \epsilon_t \qquad (6.3.3)$$

The estimation of λ and γ will be discussed later in this chapter.

(ii) time varying gain

In this case we state:

$$\underline{\Gamma}_t = \underline{\Gamma}_{t-1} + \delta\, \underline{\Gamma}_t, \text{ where}$$

$\delta\, \underline{\Gamma}_t \sim N[0,\underline{W}]$.

The model in 6.3.3 can be rewritten as:

$$\begin{cases} Y_t = \theta_t + \epsilon_t\ ; \quad \epsilon_t \sim N(0,v_t) \\ \theta_t = \lambda\, \theta_{t-1} + \gamma_t\, x_t \\ \gamma_t = \gamma_{t-1} + \delta\, \gamma_t\ ,\ \delta\, \gamma_t \sim N(0,W_\gamma) \end{cases} \qquad (6.3.4)$$

and λ is known.

(iii) stochastic gain factor.

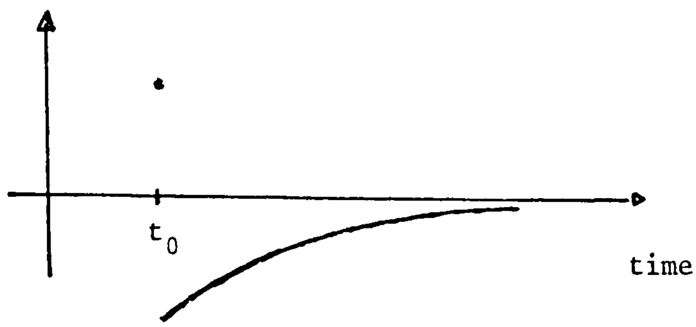
Since often the gain is stochastic we can phrase the model as:

$$\underline{\Gamma}_t \sim [\underline{\Gamma},\ V_{\underline{\Gamma}}]$$

The interpretation is that many other variables are omitted in the model and so the gain is described as a full probability distribution. The first order model 6.3.4 is rewritten as:

$$\begin{cases} Y_t = \theta_t + \epsilon_t \\ \theta_t = \lambda\, \theta_{t-1} + \gamma_t\, \alpha \\ \gamma_t = \overline{\gamma} + \delta\, \gamma_t \ ; \quad \delta\, \gamma_t \sim [0;\, v_\gamma] \end{cases}$$

It is worth mentioning one interesting example of application of these ideas in intervention analysis. For example in the champagne data analyzed in Chapter 3 $x_t = \begin{cases} 1 & \text{if } t=t_0 \\ 0 & \text{if } t\neq t_0 \end{cases}$. A component intervention model is such that the transfer response graph is like:



Of course the general case of a hierarchical model (Lindley and Smith (1972)) can be handled without any additional problem. That is to say:

$$\underline{\Gamma}_t = N[\overline{\underline{\Gamma}}_t\ \sigma^2_{\underline{\Gamma}}]$$

$$\overline{\underline{\Gamma}}_t = \overline{\underline{\Gamma}}_{t-1} + \delta\, \overline{\underline{\Gamma}}_t\ ,\ \delta\, \overline{\underline{\Gamma}}_t \sim N(0, W_{\overline{\Gamma}})$$

6.3.3 Stochastic Transfer Response Estimation.

In this section we will present a general method for on-line estimation of transfer response and the first order model and the 2nd order model are discussed. The second order model is important because this is the minimum order in which oscillatory phenomena of general periodicity occur.

(i) the observational distribution

$$(Y_t/\psi_t) \sim N(\psi_t; V_t) \tag{6.3.5}$$

where: $\psi_t = \underline{F}\ \underline{E}_t$.

$\underline{E}_t = \underline{G}(\underline{\lambda}_t)\ \underline{E}_{t-1} + \underline{\Gamma}_t\ x_t$ are the guide relationships.

In this notation $\underline{G}$ is a matrix p×p (in canonical form);

$\underline{\Gamma}_t$ is a vector p×1 of gain factors

$\underline{\lambda}_t$ is a vector r×1 representing the eigenvalues of $\underline{G}$, each one with multiplicity p_i, such that $\sum_1^r p_i = p$.

Now, let us define $\underline{\theta}_t = (\underline{\lambda}_t, \underline{\Gamma}_t, \underline{E}_{t-1})'$ and consider the guide relationship

$$\underline{\theta}_t = \underline{G}^* \begin{pmatrix} \underline{\theta}_{t-1} \\ \underline{E}_{t-1} \end{pmatrix} \tag{6.3.6}$$

where:

$$\underline{G}^* = \begin{bmatrix} I_r & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & I_p & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & I_p \end{bmatrix}$$

(ii) The joint prior distribution is obtained as:

$$\left(\begin{matrix} \psi_t \\ \underline{E}_t \\ \underline{\theta}_t \end{matrix} \middle| D_{t-1} \right) \sim N\left[\begin{pmatrix} \hat{\psi} \\ \hat{\underline{E}} \\ \hat{\underline{\theta}} \end{pmatrix}_t ;\ \begin{bmatrix} r_{11} & \underline{r}_{12} & \underline{r}_{13} \\ \underline{r}_{21} & \underline{r}_{22} & \underline{r}_{23} \\ \underline{r}_{31} & \underline{r}_{32} & \underline{r}_{33,t} \end{bmatrix} \right]$$

where:

$$\begin{cases} \hat{\psi}_t = \underline{F}\ \hat{\underline{E}}_t \\ \hat{\underline{E}}_t = E\{g(\underline{\theta}_t, x_t)/D_{t-1}\} \\ \hat{\underline{\theta}}_t = \underline{G}^*\ \underline{m}_{t-1} \end{cases}$$

$$\begin{cases} r_{11} = \underline{F}\ \underline{r}_{22}\ \underline{F}' \quad ; \qquad \underline{r}_{12} = \underline{F}\ \underline{r}_{22}\ ;\ \underline{r}_{13} = \underline{F}\ \underline{r}_{23} \\ \underline{r}_{22} = \mathrm{Var}\{g(\underline{\theta}_t, x_t)/D_{t-1}\}\ ;\ \underline{r}_{23} = \mathrm{Cov}\left(g(\underline{\theta}_t, x_t);\ \underline{\theta}_t/D_{t-1}\right); \\ r_{33,t} = \underline{B}^{-\frac{1}{2}}\ \underline{G}^*\ \underline{C}_{t-1}\ \underline{G}^{*\prime}\ \underline{B}^{-\frac{1}{2}} + \underline{W}_t\ , \quad \text{with} \end{cases}$$

$\underline{B} = \mathrm{diag}(\beta\ I_r,\ I_p,\ I_p)$, $\underline{W}_t = \mathrm{diag}(0,\ \mathrm{diag}\ (0,0\ldots0,W_r,0,0\ldots0)$

(iii) The prior to posterior evolution is calculated using the methods of Chapter 3. Let the posterior distribution for ψ_t be: $(\psi_t/D_{t-1};y_t) \sim N(m,\sigma_{11})$ which is calculated from (i) using the standard Bayesian analysis.

The break back of information is:

$$\left(\begin{matrix}\underline{\theta}_t \\ \underline{E}_t\end{matrix} \middle| D_t\right) \sim N(\underline{m}_t,\underline{C}_t), \text{ where:}$$

$$\begin{cases} \underline{m}_t = \hat{\underline{\theta}}_t + \underline{r}' \; [m-\hat{\psi}_t]/r \\ \underline{C}_t = \underline{R}_t - \underline{r}' \; r[r-\sigma_{11}]/r^2 \end{cases}$$

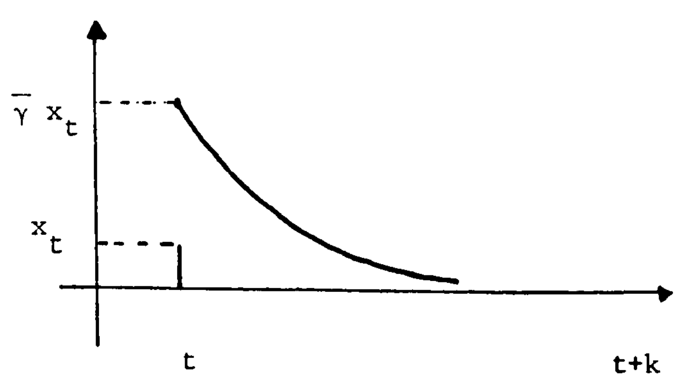
with: $\underline{r} = (\underline{r}_{12} \; \underline{r}_{13})'$; and $r = r_{11}$ and $\underline{R}_t = \begin{bmatrix} \underline{r}_{22} & \underline{r}_{23} \\ \underline{r}_{32} & \underline{r}_{33} \end{bmatrix}$

(iv) Examples:

(a) The first order model

For this particular example the impulse response function represents an exponential decay;

$$F_t(k) = E[Y_{t+k}/D_t] = \lambda^k.E[\gamma_{t+k}/D_t] \; x_t$$



The model can be described as:

$$\underline{F} = (1); \; \underline{G}^* = \begin{bmatrix} I_2 & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & 1 \end{bmatrix} \quad ,$$

$\underline{\theta}_t = (\lambda_t,\gamma_t,E_{t-1})$, with the guide relationship:

$$E_t = \lambda_t \underline{E}_{t-1} + \gamma_t \, x_t \;, \quad x_t = \begin{cases} 1 & \text{if } t=t_0 \\ 0 & \text{if } t\neq t_0 \end{cases}$$

(b) Second order model.

This is a very useful model to represent oscillatory phenomena which occur in equilibrium seeking systems involving human intervention. As pointed out before, the superposition principle can be used to model the complete system as a combination of 1st and 2nd order submodels covering a wide range of applications. Since the eigenvalues of the system evolution matrix completely describe the forecasting function (or the response pattern) we must consider three different cases, that is:

(i) Unequal Real Roots.

The transfer response form in this case is:

$$F_t(k) = [\lambda_1^k \bar{\gamma}_1 + \lambda_2^k \bar{\gamma}_2] x_t$$

which is a combination of two exponential decay. The model can be phrased as:

$$\underline{F} = (1,1) \ ; \ \underline{G}^* = \begin{bmatrix} I_4 & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{I}_2 & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{I}_2 \end{bmatrix} \ ;$$

$\underline{\theta}_t = (\lambda_{1,t}, \lambda_{2,t}, \gamma_{1,t}, \gamma_{2,t}, E_{1,t-1}, E_{2,t-1})'$ and the guide relationships:

$$\begin{cases} E_{1,t} = \lambda_{1,t} E_{1,t-1} + \gamma_{1,t} x_t \\ E_{2,t} = \lambda_{2,t} E_{2,t-1} + \gamma_{2,t} x_t \end{cases}$$

(ii) Equal Roots Case:

The form is:

$$F_t(k) = [\bar{\gamma}_1 + \frac{k}{\lambda} \bar{\gamma}_2]\lambda^k x_t$$

and the model is written as:

$$\text{guide:} \quad \underline{E}_{1,t} = \lambda_t E_{1,t-1} + E_{2,t} + \gamma_{1,t} x_t$$

$$\underline{E}_{2,t} = \lambda_t E_{2,t-1} + \gamma_{2,t} x_t$$

$$\underline{F} = (1,0) \; ; \; \underline{G}^* = \begin{bmatrix} 1 & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & I_2 & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & I_2 \end{bmatrix}$$

(iii) Finally the complex conjugate case has a forecast of the form:

$$F_t(k) = [\overline{\psi}_1 \cos(kw) + \overline{\psi} \sin(kw)]\phi^k x_t$$

and can be modelled as:

$$\text{guide:} \quad \begin{cases} E_{1,t} = \phi_t [\cos(w) E_{1,t-1} + \sin(w) E_{2,t-1}] + \gamma_{1,t} x_t \\ E_{2,t} = \phi_t[-\sin(w) E_{1,t-1} + \cos(w) E_{2,t-1}] + \gamma_{2,t} x_t \end{cases}$$

with:

$$\underline{F} = (1,1) \; , \; \underline{G}^* = \begin{bmatrix} 1 & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & I_2 & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & I_2 \end{bmatrix} \quad \text{and}$$

parameterization $\underline{\theta}_t = (\lambda_t, \gamma_{1,t}, \gamma_{2,t} \; E_{1,t-1} \; E_{2,t-1})'$. In this case $w = \frac{2\pi}{T}$ for T known.

6.4 Simulation study.

To illustrate and access the performance of the estimation method of Section 6.3 we performed a simulation study. We generated for the 1st and 2nd order transfer response models 200 sequences of 35 observations each.

The first order model is described in Section 6.3.3 and the simulation results are presented in tables 1, 2 and 3. We have simulated two different values for λ, i.e. $\lambda = .5, .9$ and in the static case the gain factor γ was fixed as 100. For the stochastic model γ was assumed to be N(100; 400). For different values of the pair (λ,γ), the cases of 1, 2 and 7 unit inputs were considered; $x_t = 1$ if $t = t_0$ and $x_t = 0$ otherwise. These unit inputs were equally spaced in the sample interval. All the results presented below are based on a very flat initial prior.

For example initially the mean value of the gain was set equal zero with a variance of 1.0 E12.

Table 1: Means and square roots of the MSE for the non-linear estimation: Simulated 1st order static transfer response model; 200 sequences of length 35 in each group.

λ	no. of inputs	Parameters			
		λ		γ	
		mean	s. error	mean	s. error
.5	1	.4985	.025	100.99	10.316
	2	.4983	.022	100.97	6.902
	7	.4984	.020	100.30	3.665
.9	1	.8992	.004	100.90	5.758
	2	.8991	.004	100.90	5.314
	7	.8994	.004	100.53	4.347

Looking at the figures in table 1 we notice that estimates are improved with the number of inputs and that systematically we under-estimate λ and over estimate γ. This is not a surprise since there is a strong correlation between λ and γ (-.6 and -.52 for 1 and 2 inputs respectively).

The results for the stochastic model are similar to the static case, and it is worth pointing out that with 7 inputs we get a really good estimate of the "true" $\bar{\gamma}$.

Table II: Mean and square root of the MSE. Non-linear estimation; simulated 2nd order transfer response ; 'stochastic'; 200 sequences of length 35 in each group.

λ	no. of inputs	Parameters			
		λ		γ	
		mean	s. error	mean	s. error
.5	1	.4985	.025	103.120	21.641
	7	.4984	.025	99.998	8.397
.9	1	.8991	.004	102.973	19.397
	7	.8993	.005	100.317	8.547

Note: For 1 input the true $\bar{\gamma}$ is equal to 102.12 and for 7 inputs it is 101.32 .

Finally in table III we have some results for the posterior variance of γ through the time. It is clear that this variance approaches a limiting value very fast and it depends upon the number of inputs. Note that the inputs occured at time 1 and 21.

Table III: Posterior variance for γ.
Simulated 1st transfer response - "static",
λ = .5; two inputs.

Time	No. of inputs	
	1	2
5	117.834	117.839
10	116.949	116.949
15	116.944	116.944
20	116.944	116.944
25	116.944	51.206
30	116.944	51.156
35	944	156

Comments: (i) The distribution of the response $\underline{\theta}$ is very asymetric;

(ii) The joint distribution of λ, γ is well approximated by a Bivariate Normal with a high negative correlation.

For the second order model we considered only the case of two distinct real roots. This model is described in example b of Section 6.3.3.

The results for 200 sequences of size 35 each are presented in Table IV. These sequences were simulated with $(\lambda_1, \lambda_2, \gamma_1, \gamma_2) = (.8, .4, 67.0, 33.0)$ and the number of inputs considered were 1, 5 and 7.

Table IV :

λ	No. of inputs	Parameters			
		λ_1	λ_2	γ_1	γ_2
.8;.4	1	.794 (.011)	.253 (.055)	54.7 (1.64)	41.6 (1.74)
	5	.799 (.011)	.345 (.057)	68.1 (6.06)	32.0 (5.72)
	7	.792 (.011)	.305 (.056)	72.3 (5.94)	28.0 (5.50)

Note: The figures inside brackets are the square root of the main square error.

In table V we find the correlation among the parameters for the case of 1 input. As we can see there is a strong negative correlation between γ_1 and γ_2, and a high negative correlation between λ_2 & γ_1. These correlations may explain the strong bias in the estimation of λ_2 and γ_2.

Table V: Posterior Correlation at the end (observation 35); 1 input

	λ_2	γ_1	γ_2
λ_1	.30	-.24	.25
λ_2		-.46	.47
γ_1			-.91

The evolution through time for the posterior variance are present in Table VI for the cases of 1 and 5 inputs. Again there is a strong evidence that these variance approaches to a limit which depends, among other things, upon the number of inputs.

Table VI: Posterior variance for $(\gamma_1\ \gamma_2)$;

2nd order transfer response - "static"

$\lambda_1 = .8,\ \lambda_2 = .4$

	No. of inputs			
Time	1		5	
	γ_1	γ_2	γ_1	γ_2
5	461.60	458.92	461.60	458.92
10	454.11	451.07	433.73	425.60
15	452.69	449.59	415.93	410.29
20	451.87	448.73	319.84	300.60
25	451.40	448.24	118.20	107.74
30	451.23	448.06	78.70	71.87
35	451.20	448.02	49.37	44.99

It is worth commenting that all these results were obtained from a very vague prior at the beginning. That is:

$$(\lambda_1\ \lambda_2\ \gamma_1\ \gamma_2)'/D_0 \sim N[(.5,\ 0.0,\ .0,\ .0)';\ \mathrm{diag}(.25,\ .25,\ 1000;\ 1000)].$$

More experience with this class of sequential estimator for transfer response is still needed. A study investigating the loss of efficiency when estimating the 2nd order TR must consider the cases of equal roots and complex roots.

Chapter 7: Long Term Forecasting

7.1 Introduction.

In this chapter we present an application of non-linear Bayesian forecasting to long-term forecasting. The evolution of a given time series is approximated by means of a growth curve of the type used in biological statistics. This approach assumes that there is some sort of natural underlying process to be followed and that variables other than time only produce fluctuations about the long term trend.

The models for long term forecasting are very important in economic forecasting. For example they provide a guide to management in analysis of investments and in other aspects of corporated forward planning, such as the estimate of the size of plants, capital outlay, etc. Although other methods are discussed in the literature (Pearce & Harrison (1972)) we will be concerned only with mathematical trend curves applied to long term forecasting.

A trend projection forecast is essentially an average estimate of the growth process of interest (for e.g. consumption or production) in a normal year or period of time. The actual process fluctuates above and below the trend line owing to the effects of variation in other variables (for e.g. the economy performance) and random fluctuations, new end uses, technological improvements, etc.

Our method can simultaniously deal with the short and long term components. For some theoretical results about the general exponential weighted regression, the paper of Harrison and Akram (1983) is referred.

It is worth pointing out that the justification for the use of a particular trend representation is primarily empirical and so a trend forecast should be supported by other methods.

In addition a lookout should be kept for major technological and sociological developments which may abruptly change the current trend. For example in economic forecasting many factors which affect supply and demand, such as, the general state of the economy, the policy of the government and other qualitative factors (which we can think of as interventions in the process), must be taken in account. In general the economic and commercial intelligence provides information about the process of interest and the Bayesian modelling is useful for incorporating these subjective informations into the forecasting system.

In section 7.2 we present the class of modified exponential curves and some of its properties. A brief discussion of the variation about trend curves is introduced in section 7.3. Finally in section 7.4 the non-linear Bayesian Model developed in Chapter 3 is applied as a method of estimation, and some examples are shown in section 7.5.

7.2 The general modified exponential family of curves.

Let us consider the family of curves defined by:

$$Y_t = \{a+b\ r^t\}^{\phi} \tag{7.2.1}$$

where a, b are real parameters, $r \epsilon (0,1)$ and ϕ is assumed known. It is clear that this growth curve has an asymptote at $Y_t = a^{\phi}$ which is approached as t increases.

Some special cases of particular interest are:

(i) The Modified exponential curve (ϕ=1)

$Y_t = a + b\ r^t$; where a, b and r are constants with $0<r<1$.

(ii) The logistic curve (ϕ=-1)

$$\frac{1}{Y_t} = a + b\ r^t;$$

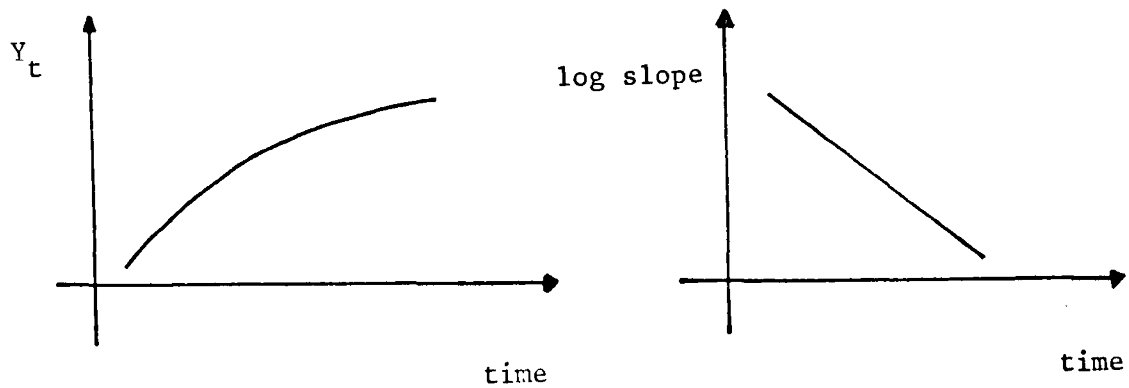
(iii) The Gompertz curve

$$\log Y_t = a + b\ r^t;$$

As pointed out by Gregg et al, it is difficult to estimate the parameter in the functional form 7.2.1. However some insight about ϕ can be obtained plotting the rate of change (slope) against the time. For example, in the simple exponential case, the log of the slope when plotted against the time gives a straight line sloping down to the right:

$$\frac{dy}{dt} = b \log r \; r^t, \; b<0$$

therefore: $$\log\left(\frac{dy}{dt}\right) = \log(b \log r) + t \log r$$



The general formula for the log of the slope can easily be obtained from 7.2.1, so:

$$\log \frac{dy}{dt} - \left(\frac{\phi-1}{\phi}\right) \log Y = \log(\phi \; b \log r) + t \log r \qquad (7.2.2)$$

If the left hand side is plotted against t, it should show a linear trend if the modified exponential fits the data. The slope characteristic can be used to assess the most suitable transformation ϕ of the original data (see Gregg et al (1964)).

Finally, it is worth pointing out that the general fitting of modified exponential must consider the variance varying on the time. In general we can state that

$$V_j = \text{var}(Y_j) \propto \alpha \; \hat{y}_j^{\beta}, \qquad \hat{y}_j = E(Y_j)$$

To a first approximation

$$V_j = V(Y_j^{1/\phi}) \simeq \phi^2 \, \hat{y}_j^{2(\phi-1)/\phi} V(y_j)$$

or

$$V_j = \phi^2 \, \alpha \, \hat{y}_j^{\beta-2(1-\phi)/\phi}$$

With $\phi=1$ we have: $V_j = \alpha \, \hat{y}_j^{\beta}$; and with $\phi = -1$, we have:

$$V_j = \iota\alpha \; \hat{y}_j^{\beta+4}.$$

7.3 Characterising the variation about trend curves.

Many different ways for characterising the variation about trend have been discussed in the literature. After an initial trend fit the variation about the trend line may be related to an independent variable, modelled as a cyclical component or respresented by a non-stationary process. A complete discussion of these procedures can be found in Harrison and Pearce (1972).

(a) Using an independent variable.

Let z_t be the variation about the trend and x_t be the variations about the trend of a relevant variable of interest. The fitted relation is of the form:

$$z_t = f(x_t), \text{ f is often a simple linear function.}$$

It follows then that the general prediction of Y_t is only benefitted if X_t is easier to forecast than Y_t. Of course these models can be viewed as examples of transfer response and they provide the basis for structural models as opposite to the conjunctural models.

(b) The cyclical representation.

A very useful way to model the variations about trend for economic variables is to include a cyclical component. The main assumption is that the control of the economy through variables like unemployment, prices etc. produce a transfer response pattern common to many economic indicators which is a fairly consistent cycle. This representation may be used to obtain a better trend estimate and a short-term protection from the effects of the cycle for forecasts in the short-term.

(c) The stochastic representation.

The idea in this case is to represent the fluctuations about the trend as an ARMA process. Since in practice any ARMA process can be well approximated by an AR(p), with p small, only this case will be considered. As in the cyclic representation we are looking for some sort of short term protection.

In the coming section we present some applications of the non-linear Bayesian model to long term forecast in which the variations about trend are represented by a AR process. Besides the usual facilities associated with the Bayesian method it permits the joint estimation of all the parameters of the model in a sequential way.

7.4 The non-linear Bayesian model applied to growth curves.

The aims of this section are to present a class of models for long term forecasting and to develop an on-line (sequential) estimation procedure. The low and high frequency can be simultaneously estimated in a sequential way.

The process trajectory is represented by one of the curves of the family described in section 7.2 and the variation about the trend curve is represented by an autoregressive process.

It is worth pointing out that some cyclical components or medium frequency can be introduced without any difficulty. First we present a simple model for trend with random shock. The on-line variance estimation is discussed. In order to complete this section the general model is discussed and some practical comments and suggestions are left out for the coming section.

7.4.1. Trend plus random shock.

Suppose that the trajectory of the process of interest is represented by one of the curves of the family in section 7.2:

$$\hat{y}_t(k) = E[Y_{t+k}|D_t]$$
$$= m_{1,t} + m_{2,t}\,\lambda^k \qquad (7.4.1)$$

where: $m_{1,t} = E[\mu_t|D_{t-1}]$; $m_{2,t} = E[\eta_t|D_{t-1}]$ and λ is assumed known.

This forecast function can be interpreted as a modified exponential as discussed before. The original observations are transformed using the desired value of ϕ as discussed in section 7.2.

The recommended NDBM $\{(\underline{F}, \underline{G}, \underline{\beta}, V)_t;\ (\underline{m}_0, \underline{C}_0)\}$ is characterized by:

$$\underline{F} = (1,1) \quad ; \qquad \underline{G} = \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} ,$$

$$\underline{\beta} = \mathrm{diag}(\beta_1\ \beta_2), \text{ with the constraint } \beta_2 < \lambda^2.$$

With this formulation, the parameters μ_t and η_t are related to the asymptote and penetration, and λ is interpreted as the penetration rate. Assuming λ known the usual Kalman filter recurrence equations can be used, and the observational variance could be estimated on-line.

Since our interest is in the case of λ unknown, the model is reformulated and the method presented in Chapter 3 is applied. The non-linear dynamic model is:

(i) observational distribution.

(a) $(Y_t|\psi_t) \sim N[\psi_t, V_t]$, where

$V_t = a\,\{E[Y_t|D_{t-1}]\}^{b_0}$ with b_0 known, and

(b) $(\psi_t|D_{t-1}) \sim N[\hat{\psi}_t, r]$

(ii) guide relationships.

(a) $\psi_t = g(\underline{\theta}_t)$, where $\underline{\theta}_t = (\mu_t, \lambda_t, \eta_{t-1})'$, and

(b) $\eta_t = \lambda_t \eta_{t-1}$ (7.4.2)

(c) the system dynamics follows the relationship:

$$\underline{\theta}_t = \underline{G} \begin{pmatrix} \underline{\theta}_{t-1} \\ \eta_{t-1} \end{pmatrix} \tag{7.4.3}$$

where $\underline{G} = \begin{bmatrix} I_2 & \underline{0} & 0 \\ \underline{0} & 0 & 1 \end{bmatrix}$.

(iii) The link between $(\underline{\theta}_t | D_{t-1})$ and $(\underline{\theta}_{t-1} | D_{t-1})$ is obtained using (7.4.3) and the prior distribution for $(\underline{\theta}_t, \eta_t)'$ is:

$$\begin{pmatrix} \underline{\theta}_t \\ \eta_t \end{pmatrix} \Bigg| D_{t-1} \sim N\left[\begin{pmatrix} \hat{\underline{\theta}}_t \\ \hat{\eta}_t \end{pmatrix} ; \begin{vmatrix} \underline{R}_{11} & \underline{R}_{12} \\ \underline{R}_{21} & R_{22} \end{vmatrix}\right]$$

where:

$$\underline{R}_{11} = \underline{B}^{-\frac{1}{2}} \quad \underline{C}_{t-1} \quad \underline{B}^{-\frac{1}{2}}$$

$$\underline{R}_{12} = \text{cov}(\underline{\theta}_t, \eta_t | D_{t-1}) \qquad \hat{\eta}_t = E(\eta_t | D_{t-1})$$

$$\underline{R}_{22} = \text{var}(\eta_t | D_{t-1}) \qquad \text{and} \quad \hat{\underline{\theta}}_t = \underline{G}\ m_{t-1}$$

Remember that $\hat{\eta}_t$, $\underline{R}_{12}$ and $\underline{R}_{22}$ are calculated using the guide relationship and the moments of a normal.

(iv) Joint prior distribution.

Let $\underline{\pi}_t = (\underline{\theta}_t, \eta_t)'$.

$$\begin{pmatrix} \psi_t \\ \underline{\pi}_t \end{pmatrix} \Bigg| D_{t-1} \sim N \begin{pmatrix} \hat{\psi}_t \\ \hat{\pi}_t \end{pmatrix} ; \begin{bmatrix} r & \underline{r} \\ \underline{r}' & \underline{R} \end{bmatrix} \tag{7.4.4}$$

where: $\hat{\psi}_t = E(\psi_t | D_{t-1})$

$$= \hat{\mu}_t + \hat{\eta}_t$$

$r = \text{var}(\mu_t + \eta_t | D_{t-1})$ and $\underline{r} = \text{cov}(\psi_t, \underline{\pi}_t | D_{t-1})$

(v) Updating procedure.

Let the posterior distribution for ψ_t be:

$(\psi_t | y_t, D_{t-1}) \sim N[m, \sigma_{11}]$, which is obtained from (i) using the standard Bayesian analysis. This information is then conveyed back to $\underline{\pi}_t$, so that:

$$(\underline{\pi}_t | D_t) \sim N[\underline{m}_t, \underline{C}_t]$$

where:

$$\underline{m}_t = \underline{\hat{\theta}}_t + \underline{r}' \, [\underline{m} - \hat{\psi}_t]/r$$

$$\underline{C}_t = \underline{R}_t - \underline{r}' \, \underline{r}[r - \sigma_{11}]/r^2 \qquad (7.4.5)$$

(vi) k-steps ahead forecasting.

The predictive distribution for ψ_{t+k} and Y_{t+k}, $k>0$ can be obtained in the following steps:

(a) $$(\underline{\pi}_{t+k} | D_t) \sim N[\underline{m}_t(k), \underline{R}_t(k)] \qquad (7.4.6)$$

where: $\underline{m}_t(k) = \begin{pmatrix} \underline{\hat{\theta}}_t(k) \\ \hat{\eta}_t(k) \end{pmatrix}$, $\underline{\hat{\theta}}_t(k) = \underline{G} \, \underline{m}_t(k-1)$, $\hat{\eta}_t(k) = E(\eta_t(k)/D_t)$

and:

$$\underline{R}_t(k) = \begin{bmatrix} \underline{R}_{11,t}(k) & \underline{R}_{12,t}(k) \\ \underline{R}_{21,t}(k) & \underline{R}_{22,t}(k) \end{bmatrix} \text{, with}$$

$$\underline{R}_{11,t}(k) = \underline{B}^{-\frac{1}{2}} \underline{R}_{11,t}(k-1) \, \underline{B}^{-\frac{1}{2}}$$

$$R_{22,t}(k) = \mathrm{var}\left(\eta_t(k) | D_t\right)$$

$$\underline{R}_{12,t}(k) = \mathrm{cov}\left(\eta_t(k), \underline{\theta}_t(k) | D_t\right)$$

Although this set for $\underline{R}_{11,t}(k)$ is not coherent it is useful for practical purposes as commented in chapter 3. From the practical point of view the variance, covariances and expected value of the non-linear guide relationship can be calculated by a first order approximation.

(b) The prior distribution for ψ_{t+k} is:

$$(\psi_{t+k}|D_t) \sim N\left(\hat{\psi}_t(k),\ r_t(k)\right)\ \forall k>0 \tag{7.4.7}$$

where:

$$\psi_t(0) = \hat{\psi}_t \text{ and } r_t(0) = r$$

(c) Finally the predictive distribution is

$$p(Y_{t+k}|D_t) \sim N[\hat{y}_t(k),\ \hat{Y}_t(k)] \quad \text{where:}$$

$$\hat{y}_t(k) = E\{E[Y_{t+k}|\psi_{t+k}]|D_t\}$$

$$= E[\psi_{t+k}|D_t] = \hat{\psi}_t(k),\ k>0$$

$$\hat{Y}_t(k) = r[E(Y_{t+k}|\psi_{t+k})|D_t] + E[var(Y_{t+k}|\psi_{t+k})|D_t]$$

$$= r_t(k) + V_{t+k}\ ;\ k>0.$$

7.4.2 The general model with autoregression disturbances.

Using the principle of superposition we will add to the basic trend model discussed before an AR(p) model to describe the disturbances.

The recommended practical extended DLM in this case is:

$$\begin{cases} y_t = \underline{F}\ \underline{\theta}_t \\ \underline{\theta}_t = \begin{bmatrix} J_2 & 0 \\ 0 & J_p \end{bmatrix} \underline{\theta}_{t-1} + \underline{w} \end{cases} \tag{7.4.4}$$

where: . J_p is a p square Jordan block with $j_{ii} = \phi_i$ i=1,...p and $j_{i,i+1}=1$, i=1,...p-1 and $j_{ik}=0$ otherwise.

. $\underline{F} = (\underline{F}_1\underline{F}_2)$; $\underline{F}_1 = (1,0)$ and $\underline{F}_2 = (1,0...0)$

$$\underline{w} \sim N(0,\underline{W}) \text{ with}$$

. $\underline{B} = \text{diag}\{\beta_1\beta_2;\ I_p\}$ and $\underline{W} = \text{diag}(0\ldots0,V,0\ldots0)$

If λ and ϕ's are known a straightforward application of the KF recurrence provides the predictive distributions.

In order to learn about λ and ϕ's the non-linear Bayesian model is used.

Definitions and notation.

(1) Let $\underline{\theta}_t$ be $\underline{\theta}_t = \begin{bmatrix} \underline{\theta}_t^{(1)} \\ \underline{\theta}_t^{(2)} \end{bmatrix}$, $\underline{\theta}_t^{(1)} = (\mu_t,\lambda_t,\eta_{t-1})'$ and

and $\underline{\theta}_t^{(2)} = (\underline{\phi}_t,\ \underline{\alpha}_{t-1})'$, where:

$\underline{\phi}_t$ is a p×1 vector of autoregressive coefficients, and $\underline{\alpha}_t$ is a p×1 vector

(2) Define the (3p+3) × 1 vector $\underline{\pi}_t$ as

$$\underline{\pi}_t = \left(\underline{\theta}_t^{(1)},\ \eta_t,\ \underline{\theta}_t^{(2)},\ \underline{\alpha}_t\right)$$

(3) Let $\underline{G}$ be $\underline{G} = \text{diag}\ (\underline{G}_1,\ \underline{G}_2)$ where

$$\underline{G}_1 = \begin{bmatrix} I_2 & \underline{0} & 0 \\ \underline{0} & 0 & 1 \end{bmatrix} \quad \text{and } \underline{G}_2 = \begin{bmatrix} \underline{I}_p & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & I_p \end{bmatrix}$$

(4) Guide relationships.

(a) Parameter evolution:

$$\underline{\theta}_t = \underline{G}\ \underline{\pi}_{t-1}$$

(b) $\psi_t = \mu_t + \eta_t + \alpha_{1,t}$

(c) $\eta_t = \lambda_t\ \eta_{t-1}$ and $\alpha_{i,t} = \phi_{i,t}\ \alpha_{i,t-1} + \omega_i$, where

$$\omega_i = \begin{cases} \omega & i=1 \\ 0 & i\neq1 \end{cases} \quad \text{and } \omega \sim N(0,V_t)$$

Prior distribution for $\underline{\theta}_t$.

Suppose that $(\underline{\pi}_{t-1}|D_{t-1})$ is $N(\underline{m}_{t-1},\underline{C}_{t-1})$ and use 4(a) to get

$$(\underline{\theta}_t | D_{t-1}) \sim N(\hat{\underline{\theta}}_t, \underline{R}_{11,t}) \tag{7.4.5}$$

where:

$$\begin{cases} \hat{\underline{\theta}}_t = \underline{m}_{t-1} \\ \underline{R}_{11,t} = \underline{B}^{-\frac{1}{2}} \; \underline{C}_{t-1} \; B^{-\frac{1}{2}} + W \end{cases}$$

with:

$$\underline{B} = \mathrm{diag}(\beta_{ii}) \text{ with } \beta_{ii}=1, \text{ if } i=4+p$$

$$\underline{W} = \mathrm{diag} \text{ with } W_{ii} = \begin{cases} W & \text{if } i=4+p \\ 0 & i \neq 4+p \end{cases}$$

From a practical view point it is worth considering the modified discounting as in Chapter 3.

Joint Prior distribution for $\underline{\psi}_t$, $\underline{\pi}_t$

$$\left(\begin{matrix} \psi_t \\ \underline{\pi}_t \end{matrix} \middle| D_{t-1} \right) \sim \left[\begin{pmatrix} \hat{\psi} \\ \hat{\pi}_t \end{pmatrix} ; \begin{bmatrix} r & \underline{r} \\ \underline{r}' & \underline{r} \end{bmatrix} \right]$$

where:

$\hat{\pi}$ and $\underline{R}$ are calculated using the guide relationship and the general moments for a Normal.

The updating.

Using the non-linear method with some adaptation we can write:

$$\begin{cases} \underline{m}_t = \underline{\hat{\pi}} + \underline{r}(y_t - \hat{y}_t)/r \\ \underline{C}_t = \underline{r}_t - \underline{r}' \; \underline{r}/r. \end{cases}$$

7.5 Applications.

7.5.1 General.

The method developed in the previous section is applied to two data sets. In the first example we show an application to some annual data and in the second to a quarterly data set. The components of the later model are trend, seasonal and an autoregressive term, which are jointly estimated on line.

From the discussions in section 7.2 we get some insight about the initial setting of the parameters for the trend component. Usually we have some initial information about the asymptotic level. For example it depends upon the size of the economic system as a whole and so it is not difficult to set its mean value. To estimate the initial mean value for λ consider N as the time for the penetration to reach a half of the total. Then a good working approximation to the value of λ is:

$$\lambda \simeq (3N-1)/(3N+1)$$

or precisely:

$$\lambda = \left\{\frac{a}{b} - \tfrac{1}{2}\right\}^{1/N}$$

For example if N=50 then λ=.986. Finally the total penetration at the beginning can be roughly set as the difference between the asymptote and the first observation.

With respect to the AR component we suggest an inspection of the residuals after the fitting of the trend to estimate the order of the AR component as well as the initial value of its components.

Generally speaking the information at the beginning can be viewed as representing a vague initial prior.

Finally the discount factors must be defined through the following rules:

(i) a high discount for the asymptote reflecting our views that it evolves smoothly through time. In general we suggest putting $\beta = .99$;

(ii) the same sort of arguments are valid for the penetration rate λ;

(iii) for the current penetration level we generally follow the condition that $\beta<|\lambda|^2$ (see Harrison and Akram (1982)), and a reasonable working rule is $\beta_t \approx .95\ \lambda_t^2\ \forall\ t=1,2,\ldots T$; and,

(iv) if we want to learn about the AR pattern and its coefficients we must set the corresponding β's equal 1.

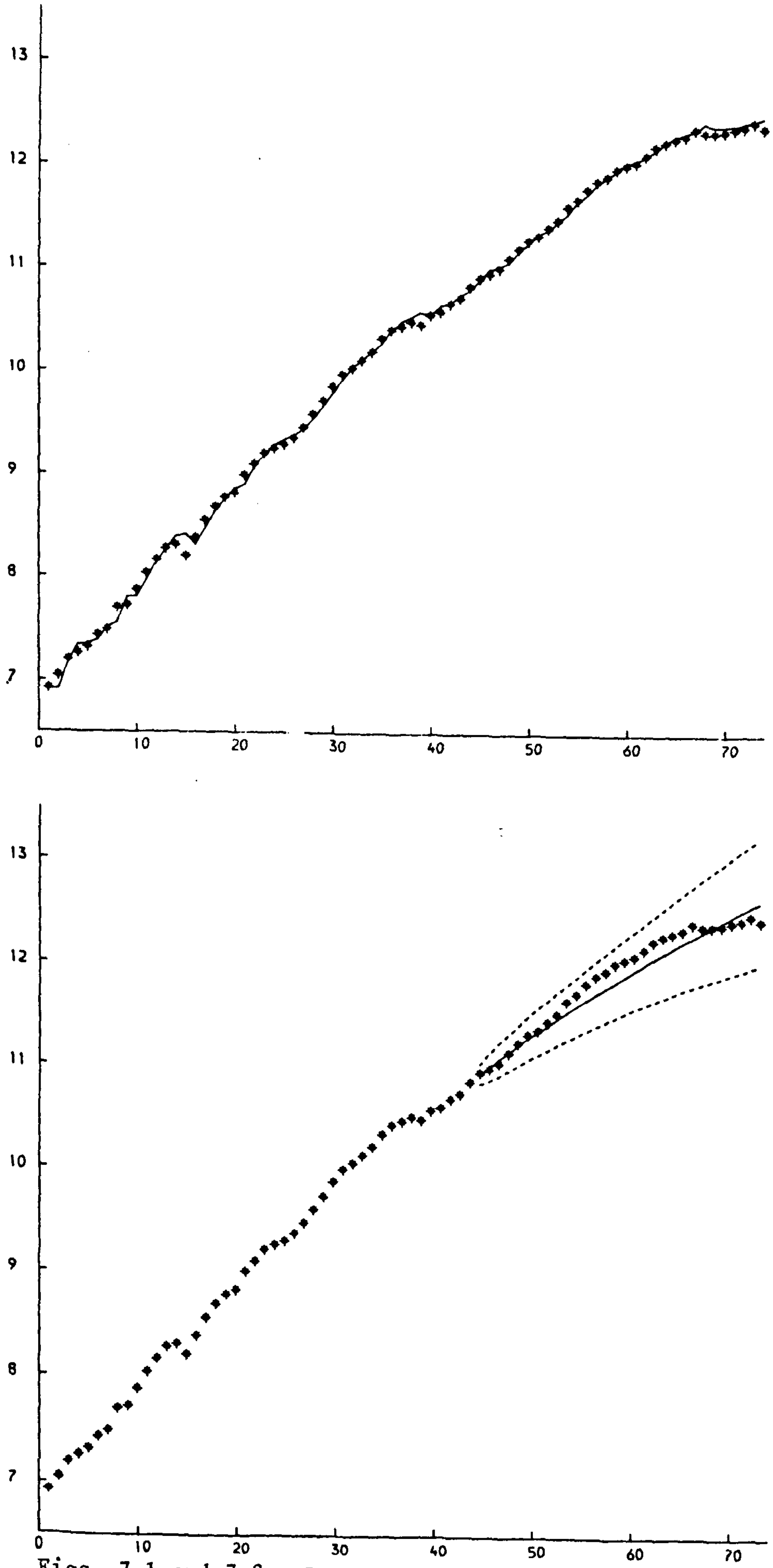
7.5.2 Examples.

Example 1. Electricity data.

The electricity data given in appendix 7.1 shows the annual available electricity in U.K. from 1907 to 1980. The model fitted to this data uses a Gompertz growth curve to represent the long-term trend plus an autoregressive order 1 to cover the short term variations.

A model with only the trend component was initially fitted to the data and the autocorrelation function calculated in order to estimate the order of the autoregressive parameter within a 95% interval (.50, .90).

The one step ahead mean point forecasts are plotted in fig. 7.1 and give an indication of the goodness of the fit. In figure 7.2 we present the trend projection for the next periods ahead made using the first 45 observation. The confidence bands associated with the long-term component projection.



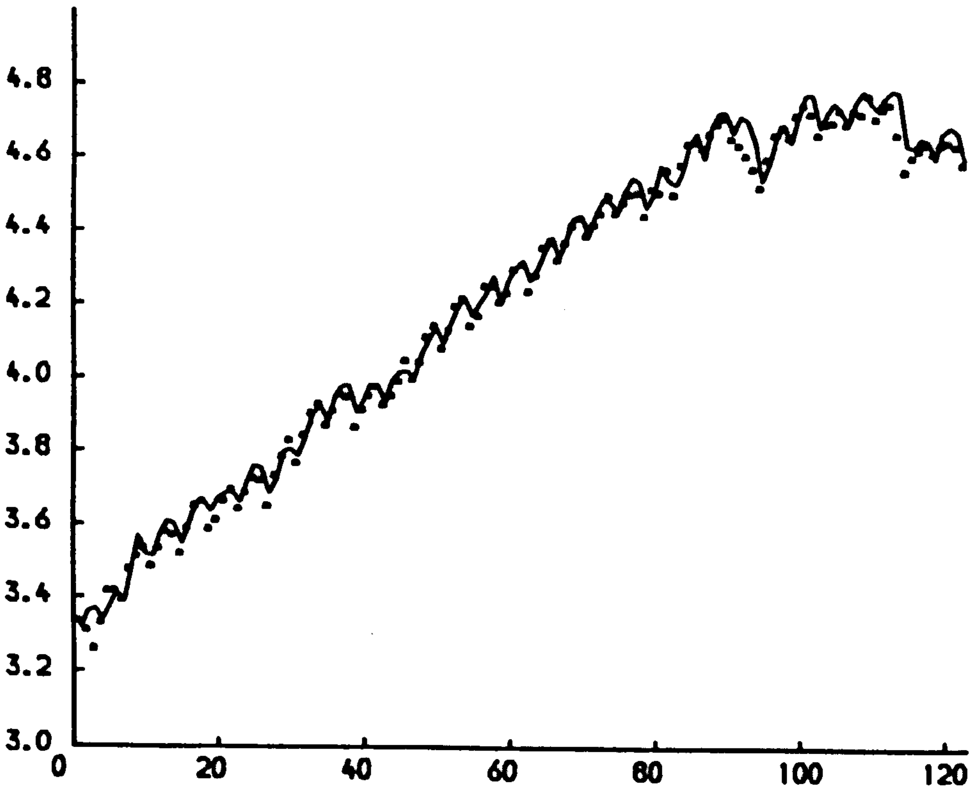
Figs. 7.1 and 7.2. Log of availability electricity (*) and one-step-ahead forecast with confidence bands (ϕ = .7, β = .995)

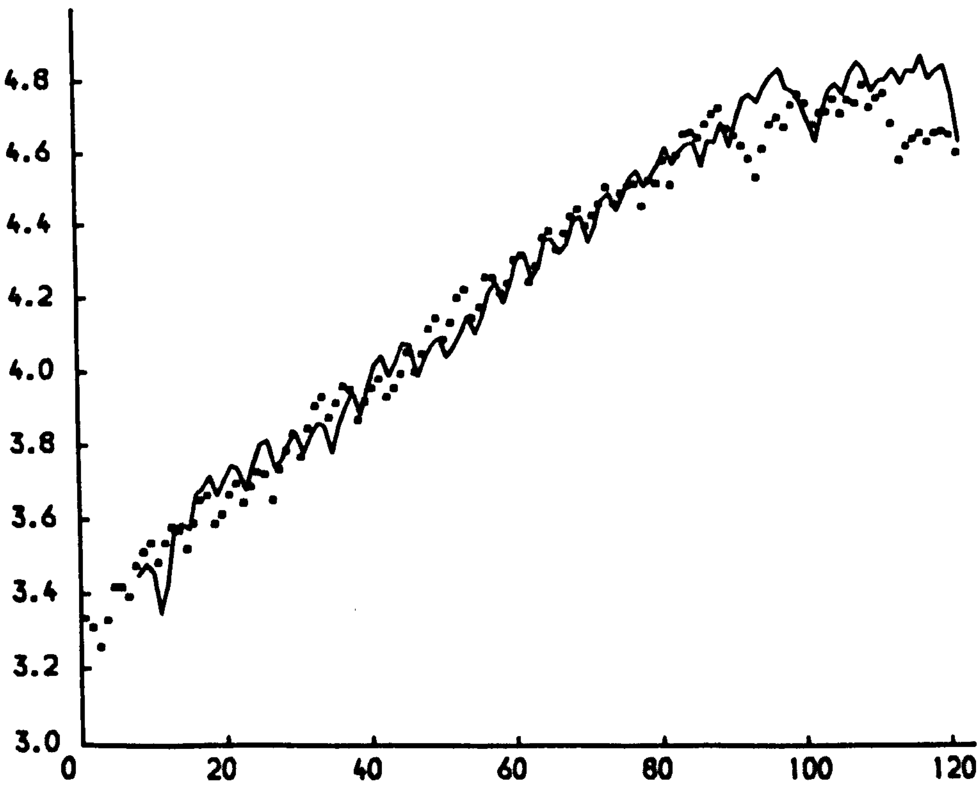
Example 2. Chemical index.

These 123 observations listed in appendix 7.2 correspond to the quarter index for industrial production for the chemical industries in U.K. from 195 to 1982. A model involving a trend, a seasonal component and a AR(1) process is used to represent the data.

The specification made at the beginning corresponds to a very weak prior setting with no seasonal pattern and an AR parameter within the 95% interval (.65, 1.0). The rate of penetration was set as .985 with a standard error of .01. The residual variance was assumed known and equal to .0001.

In fig. 7.3 we can see the one-step ahead mean forecasting and in fig. 7.4 the 8-steps ahead forecasting (or 2 years-ahead).





Figs. 7.3 and 7.4 . Chemical index data (*) one month-ahead and eight months ahead forecast (ϕ = .80, β = .995)

Appendix 7.1: Availability of Electricity (G.B.).

YEAR		YEAR	
1907	1.015	1944	36.585
1908	1.153	1945	35.706
1909	1.337	1946	39.178
1910	1.426	1947	40.369
1911	1.509	1948	43,815
1912	1.684	1949	46.241
1913	1.78	1950	51.877
1914	2.189	1951	56.658
1915	2.247	1952	58.816
1916	2.629	1953	61.955
1917	3.08	1954	68.921
1918	3.502	1955	75.648
1919	3.931	1956	82.191
1920	4.065	1957	85.999
1921	3.664	1958	93.441
1922	4.376	1959	100.572
1923	5.177	1960	114.417
1924	5.899	1961	122.823
1925	6.492	1962	136.015
1926	6.788	1963	147.102
1927	8.102	1964	154.534
1928	8.979	1965	166.051
1929	9.979	1966	171.99
1930	10.475	1967	177.309
1931	10.913	1968	190.166
1932	11.701	1969	202.593
1933	12.964	1970	211.461
1934	14.769	1971	218.053
1935	16.817	1972	224.531
1936	19.404	1973	239.651
1937	21.841	1974	231.726
1938	23.338	1975	232.431
1939	25.189	1976	234.458
1940	27.327	1977	241.174
1941	31.01	1978	246.642
1942	33.841	1979	258.045
1943	35.096	1980	246.053

UNIT: 1000 GWH

SOURCE: Digest of Energy Statistics (1934-1982)

Ministry of Power, U.K.

Appendix 7.2: U.K. Chemical Industries Quarterly Index.

YEAR				
1952	28.60	27.90	26.50	28.50
1953	31.10	31.10	30.30	33.00
1954	34.20	35.00	33.30	35.00
1955	36.60	36.30	34.50	37.00
1956	39.30	39.80	36.90	37.80
1957	39.90	41.10	39.00	40.80
1958	42.40	42.10	39.30	42.70
1959	45.00	47.00	44.20	47.70
1960	50.70	51.90	49.10	51.10
1961	53.40	52.90	48.80	51.20
1962	53.20	54.40	51.90	58.30
1963	55.30	58.50	55.70	58.30
1964	62.40	64.30	60.60	63.50
1965	67.90	69.40	64.40	66.20
1966	71.80	71.70	68.80	70.70
1967	75.30	76.30	71.00	74.10
1968	80.10	81.60	77.50	81.10
1969	84.90	86.70	82.80	85.20
1970	88.00	92.20	88.80	90.80
1971	92.60	93.00	87.40	94.10
1972	93.20	99.30	92.80	100.60
1973	106.70	107.20	105.20	109.70
1974	112.60	114.60	108.20	106.20
1975	103.20	99.70	94.60	102.40
1976	109.30	111.50	108.60	115.40
1977	118.70	115.80	109.30	112.80
1978	113.20	117.20	112.70	117.00
1979	115.90	121.90	114.50	117.40
1980	119.00	109.60	99.10	103.00
1981	104.90	106.60	104.10	106.60
1982	107.30	106.20	101.10	-

BASE: 1975 = 100

SOURCE: I.C.I. - U.K.

Chapter 8: Summary and Conclusions

8.1 Introduction

The purpose of this dissertation has been to develop dynamic Bayesian models for application in non-linear, non-normal time series and transfer response problems providing extension of the standard DLM. The analysis use the conjugate prior to the posterior distribution for the exponential family parameters, which leads to the calculation of predictive distributions in closed standard form.

The flexibility of the dynamic Bayesian linear models of Harrison-Stevens (1976) as components of structured forecasting systems and their potential as aids to interpreting the observed behaviour of real life processes, which make them suitable for applications in different fields, was kept by our proposed method. The standard normal DLM supposes that Y_t is normally distributed with mean $Y_t = E(Y_t|\underline{\theta}_t) = \underline{F}\ \underline{\theta}_t$ and known variance V_t, where $\underline{\theta}_t$ is the underlying parameters vector, whose evolution in time determines the structure of the model. This linearity assumption has been substituted by a non-linear guide relationship. Although transformations of the data enable the normal DLM to provide a useful approximation to a more complex structure, we have pointed out that these transformations cause conflict between the requirements of linearity, constant variance and normality (or, at least symmetry). It is clear that the use of the non-linear structure and the appropriate probability model avoid these problems.

A wider class of non-linear guide relationship, relating to the underlying parameters and the parameter in the exponential family, have been described in this work, although only the particular product form have been extensively used. These guide relationships, in form of sum of products of the state parameters, have enabled us to develop applications for seasonal multiplicative models, estimation of transfer response and long term models. Some more investigation is necessary to validate the method for general non-linear guide relationships.

8.2 Model Summary and Applications

The full system of recursion is summarised here :

(i) observational model

$p(y_t|\psi_t,\phi) \propto \exp[\phi\{y_t \psi_t - a(\psi_t)\}] \quad b(g,\phi)$

(ii) guide relationship

$\psi_t = g(\underline{\theta}_t) + \delta , \quad \delta \sim [o,\sigma^2]$

(iii) prior

(a) $(\psi_t|D_{t-1}) \sim CP[\alpha_t,\beta_t]$

(b) $(\hat{\underline{\theta}}_t|D_{t-1}) \sim [\hat{\underline{\theta}}_t,\underline{R}_t]$, where

CP is a conjugate prior and $(\underline{\theta}_t|D_{t-1})$ is partially specified.

The guide relationship is used to determine α_t and β_t

That is to say, linking the mean and variance of $\underline{\theta}_t$ and ψ_t :

$\hat{\psi}_t = E(\psi_t|D_{t-1}) = E[g(\underline{\theta}_t)|D_{t-1}]$

$r = V(\psi_t|D_{t-1}) = V[g(\underline{\theta}_t)|D_{t-1}]$

(iv) posterior

$(\psi_t|D_t) \sim CP[\alpha_t + \phi\, y_t, \beta_t + \phi]$

and so let $m = E(\psi_t|D_t)$ and $\sigma_{,,} = V(\psi_t|D_t)$ be obtained as function of

α_t, β_t, ϕ and Y_t.

(v) underlying parameters vector update

$(\underline{\theta}_t/D_t) \sim [\underline{m}_t, \underline{C}_t]$

$\underline{m}_t = \hat{\underline{\theta}}_t + \underline{r}_t\,[m-\psi_t]|r$

$\underline{C}_t = \underline{R}_t - \underline{r}'_t\,\underline{r}_t\,[r-\sigma_{11}]|r^2$

where these recurrence relations were obtained from the application of the linear Bayes estimation (Hartigan (1969)).

(vi) underlying parameters vector evolution

$$\underline{\hat{\theta}}_t = \underline{G}\,\underline{m}_{t-1}$$
$$\underline{R}_t = \underline{B}^{-\frac{1}{2}}\ \underline{G}\ \underline{C}_{t-1}\ \underline{G}'\ \underline{B}_t^{-\frac{1}{2}} \qquad \text{, where}$$

$\underline{B} = \text{diag}\ \{b_1\ b_2\ \ldots\ b_p\}$,$o<b_i \le 1$ and $\underline{G}$ is now an pxp matrix.

Application to various members of the exponential family have been presented with special reference to the beta-binomial (advertising example in chap. 5), the lognormal seasonal growth and the normal multiplicative seasonal growth model of chapter 3. For special examples in the context of dynamic generalized linear models the paper of West, Harrison and Migon (1984) is referred. It is worth pointing out that the method was used for estimation of normal transfer response structures and for the estimation of low and high frequencies in models for long term forecasting.

The performance of the estimation methods was extensively assessed through many applications to real and artificial data, some of which were reported in this thesis as examples of the models.

Finally we should stress the performance of the method in the application described in chapter 4 which is reported in Migon and Harrison (1983). The models developed for the Millward Brown Marketing Research Company and the programmes we have implemented in their computers have been working sucessfully for more than two years.

8.3 Limitations and directions for Further Research

1. The linear Bayes procedure is an approximate method which kept the analysis tractable in our case. In particular it provides updating recurrenas that coincide with the Kalman filter in normal case as we have shown in chapter 3. Some more investigation is necessary for an evaluation of these approximations in other cases.

2. Multiprocess mulitplacative seasonal growth model need to be developed. In real life applications the models must be able of automatically accommodating outliers or adapting to structural changes in the observed series.

3. The models relating TV advertising and awareness can be extended in different directions as pointed out by Broadbent (1979). For example we could analyse separately the effects of different sorts of time bought, as on peak and late off peak; we could consider the effect of advertising for other competitors etc.

4. More experience with transfer response is needed both from a methodological and a practical point of view. The simulation results in chapter 6 would be extended specially for second order models. From a practical point of view we would intend to do some econometric application with real data.

5. One shortcoming in the long term forecasting models of chapter 7 is concerned with the variance estimation. We believe that the model can be rewrite in a suitable form which permits the sequential estimation of the observational variance.

6. A more ambitious research is related to the extension of these univariate methods to multivariate forecasting, and perhaps to simultaneous equations econometric.

REFERENCES

ABRAHAM, B. (1980) - Intervention Analysis and Multiple Time Series. Biometrika, 67, 1, pp. 73-8.

AITCHISON, J. and DUNSMORE, I.R. (1975) - Statistical Prediction Analysis. Cambridge: University Press, London.

AKAIKE, H. (1974) - Stochastic Theory of Minimal Realization. I.EEE Transactions on Automatic Control, Dec. 1974, pp. 667-74.

ALTHAM, P.M.E. (1978) - Two Generalizations of the Binomial Distribution. Appl. Statist., 27, No. 2, pp. 162-167.

AMEEN, J.R.M. and HARRISON, P.J. (1983) - Discount Weighted Estimation. To be published in Journal of Forecasting.

AMEEN, J.R.M. and HARRISON, P.J. (1983) - Discount Bayesian Multiprocess Modelling with Cusums. Warwick Research Report No. 34.

AMEEN, J.R.M. and HARRISON, P.J. (1983) - Normal Discount Bayesian Models. Warwick Research Report No. 30.

ANDERSON, T.W. (1958) - Introduction to Multivariate Statistical Analysis - John Wiley and Sons - New York.

ÅSTRÖM, K.J. and EYKHOFF, P. (1971) - System Identification - A Survey Automatica - Vol. 7, pp. 123-62.

ATKINS, M.S. (1979) - A Case Study on the use or Intervention Analysis Applied to Traffic Accidents. J. Opl. Res. Soc., Vol. 30, 7 pp. 651-59.

AZZALINI, A. (1983) - Approximate Filtering of Parameter Driven Process. J. Time Series Analysis 3, 219-224.

BAKER, R.J. and NELDER, J.A. (1978) - Glim Release 3. N.A.G. - Oxford.

BARD, D.R. and SLEZAK, N.L. (1972) - A Comparison of Multivariate Normal Generators. Communications of the ACM, Vol. 15, No. 12.

BHATTACHARYYA, M.N. (1982) - Lydia - Pinkham Data Remodelled.Journal of Time Series Analysis, Vol. 3, No. 2, pp. 81-107.

BOX, G.E.P. (1980) - Sampling and Bayes' Inference in Scientific Modelling and Robustness. J.R. Statist. Sol. A, 143, Part 4, pp 383-430.

BOX, G. and JENKINS, G.M. (1970) - Time Series Analysis Forecasting and Control - San Francisco: Holden Day (revised edition published 1976).

BOX, G.E.P. and McGREGOR, J.F. (1974) - The Analysis of the Closed Loop Dynamic Stochastic Systems. Technometrics, Vol. 16, No. 3.

BOX, G.E.P. and MULLER, M.E. (1958) - A Note on the Generation of Random Normal Deviates.

BOX, G. and TIAO, G.C. (1973) - Bayesian Inference in Statistical Analysis. Addison-Wesley Publishing Co.

BOX, G.E.P. and TIAO, G.C. (1975) - Intervention Analysis with Application Economic and Environmental Probability. J. American Statistical Asso., Vol. 70.

BROADBENTS, S. (1979) - One Way T.V. Advertisement Work. Journal of the Marketing Research Society, Vol. 21, No. 3, pp. 139-165.

BROWN, G.C. and RUTEMILLER, H.C. (1977) - Means and Variances of Stochastic Vector Products with Applications to Random Linear Models. Management Science, Vol. 24, No. 2.

CAINES, P.E., SETHI, S.P. and BROTHERTON, T.W. (1977) - Impulse Response Identification and Causality Detection for the Lydia - Pinkham Data. Annals of Economic and Social Measurement, 6/2, pp. 147-163.

CANTARELIS, N. and JOHNSTON, F.R. (1983) - On-line Variance Estimation for the Steady State Bayesian Forecasting Model - J. Time Series Analysis - Vol. 3, No. 4, pp. 225-234.

CHADDHA, R.L. and CHITCODEKAR, S.S. (1971) - A "Generalization" of the Logistic Curves Long-Range Forecasting (1966 - 1981) of Residence Telephones. Management Science, Vol. 2, No. 2.

CHATFIELD, C. and PROTHERO, D.L. (1973) - Box-Jenkins Seasonal Forecasting: Problems in a Case-Study. J.R. Statist. Soc. A., 136. Part 3, p. 295.

CHATFIELD, C. (1977) - Some Recent Developments in Time Series Analysis - J.R. Statist. Soc. A, 140, Part 4, pp. 492-510.

CHATFIELD, C. (1980) - The Analysis of Time Series - an Introduction (2nd Edition) Chapman and Hall - London.

CLARKE, D.G. (1976) - Econometric Measurement of the Duration of Advertising Effect on Sales. Journal of Marketing Research, Vol. 13, 345-57.

COLMAN, S. and BROWN, C. (1983) - Advertising Tracking Studies and Sales Effects. Journal of the Market Research Society, Vol. 25, No. 2, pp. 165-183.

COOLEY, T.F. and PRESCOTT, E.C. (1976) - Estimation in the Presence of Stochastic Parameter Variation. Econometrica, Vol. 14, No. 1.

COOPERSMITH, L.W. (1983) - Forecasting Time Series which are Inherently Discontinuous. Journal of Forecasting, Vol. 2, pp. 225-235.

COX, R.D. and HINKLEY, D.V. (1974) - Theoretical Statistics, Chapman and Hall.

DE GROOT, M.H. (1970) - Optimal Statistical Decisions, McGraw-Hill, New York.

DURBIN, J. and MURPHY, M.J. (1975) - Seasonal Adjustment Based on a Mixed Addition Multiplicative Model. J.R. Statist. Soc. A, 138, Part 3, p. 385.

EDWARDS, A.W.F. (1972) - Likelihood. Cambridge: University Press.

ERICKSON, G.M. (1981) - Using Ridge Regression to Estimate Directly Lagged Effects in Marketing. Journal of the American Statistical Association, Vol. 76, No. 376, Applications Section.

FIELDSTEIN, M.S. (1971) - The Error of Forecast in Econometric Models when the Forecast Period Exogenous Variables are Stochastic. Econometrica, Vol. 39, No. 1.

FERGUSON, T.S. (1967) - Mathematical Statistics - A Decision Theoretical Approach Academic Press.

FILDES, R. (1983) - An Evaluation of Bayesian Forecasting - Journal of Forecasting, Vol. 2, pp. 137-50.

GREGG, J.R. et al (1964) - Mathematical Trend Curves: An Aid to Forecasting. Oliver & Boyd.

GILCHRIST, W. (1976) - Statistical Forecasting. John Wiley, New York.

GODOLPHIN, E.J. and HARRISON, P.J. (1975) - Equivalence Theorems for Polynomial Projecting Predictors. JRSS. B. 29, 355-369.

GOLDSTEIN, N. (1976) - Bayesian Analysis of Regression Problems. Biometrika 63, pp. 51-68.

GRAHAM, J.W. (1981) - Measuring Advertising's Effect on Sales. Notes on a presentation to be given at the Operational Research Society Conference at the University of Sussex.

HANSSENS, D.M. (1980) - Market Response, Competitive Behaviour and Time-Series Analysis. Journal of Marketing Research, Vol. 17, 470-85.

HARRISON, P.J. (1965) - Short-Term Forecasting. Appl. Statist., Vol. XIV, Nos. 2 and 3.

HARRISON, P.J. (1967) - Exponential Smoothing and Short-Term Sales Forecasting. Management Science, Vol. 13, No. 11.

HARRISON, P.J. and AKRAM, M. (1982) - Generalized Exponentially Weighted Regression and Parsimonious Dynamic Linear Modelling. Warwick Research Report No. 24.

HARRISON, P.J. and DAVIES, O.L. (1964) - The use of Cumulative sum (CUSUM) Techniques for the Control of Routine Forecasts of Product Demand. Operational Research Society of America 12, pp. 325-333.

HARRISON, P.J. and JOHNSTON, F.R. (1983) - A Regression Method with non Stationary Parameters. Warwick Papers in Industry, Business and Administration, No. 13.

HARRISON, P.J. and PEARCE, S.F. (1972) - The use of Trend Curves as an aid to Market Forecasting. Industrial Marketing Management, 2, pp. 149-168.

HARRISON, P.J. and STEVENS, C.F. (1971) - A Bayesian Approach for Short-Term Forecasting. Operational Research Quarterly (1971) - 22, No. 4, 341-362.

HARRISON, P.J. and STEVENS, C.F. (1976) - Bayesian Forecasting. J.R. Statist. Soc. B, 38 pp. 205-247.

HARVEY, A.C. (1981) - Time Series Models, P. Allan Publishers Ltd. England.

HELMER, R.M. and JOHANSSON, J.K. (1977) - An Exposition of the Box-Jenkins Transfer Function Analysis with an Application to the Advertising Sales Relationship. Journal of Marketing Research, Vol. 14, pp. 227-239.

JOHNSTON, F.R. and HARRISON, P.J. (1983) - Discount Weighted Moving Averages, Warwick Papers in Industry, Business and Administration, No. 14, Sept. '83.

Harrison P. J., Leonard T. and Gazzard T. N. (1977) - An Application of Multivariate Hierarchial Forecasting - Warwick Research Rep.no.15.

Hartigan J. A. (1969) - Linear Bayesian Methods - J.R.Stat.Soc.B,31,446-454

Kalman R. E. (1963) - New mthods in Wiener Filtering. Theory. Ed. Bogdanoff and Kozin F. - Wiley-N. York.

KENNY, P.B. and DURBIN, J. (1982) - Local Trend Estimation and Seasonal Adjustment of Economic and Social Time-Series. J.R. Statist. Soc. A, 145, Part 1, pp. 1-44.

LEDOLTER, J. (1979) - A Recursive Approach to Parameter Estimation in Regression and Time Series Models. Commun. Statist. - Theor. Meth., A8(12) 1227-1245.

LEONARD, T. and HARRISON, P.J. (19) - Bayesian Updating for the Steady State Kalman Filter. Warwick Research Report No. 15.

LINDLEY, D.V. (1965) - Introduction to Probability and Statistics from a Bayesian Viewpoint (2 Vols.). Cambridge: University Press.

LINDLEY, D.V. (1978) - The Bayesian Approach to Statistics. Scandinavian Journal of Statistics. Vol. 5, pp. 1-26.

LINDLEY, D.V. (1976) - A Class of Utility Functions. The Annals of Statistics, Vol. 14, No. 1, pp. 1-10.

LINDLEY, D.V. (1978) - Bayesian Statistics - A Review. Siam, Pennsylvania.

LINDLEY, D.V. and SMITH, A.F.M. (1972) - Bayes Estimate for the Linear Model. J.R. Statist. Soc. B, 34, pp. 1-39.

MAHAJAN, V. and BRETSCHNEIDER, S.J. and BRADFORD, J.W. (1980) - Feedback Approaches to Modelling Structural Shifts in Market Response. Journal of Marketing, Vol. 44, pp. 71-80.

MAKRIDAKIS, S. and WHEELWRIGHT, S.C. (1977). Interactive Forecasting. Holden-Day - San Francisco.

MAKRIDAKIS, S. and WHEELWRIGHT, S.C. (19) - Adaptive Filtering: An Integrated Autoregressive/Moving Average Filter for Time Series Forecasting. Operational Research Quarterly, Vol. 28, No. 2, ii.

MAR-MOLINERO, C. (1980) - Tractors in Spain: A Logistic Analysis. J. Opl. Res. Soc., Vol. 31, pp. 141-152.

MEHRA, R.K. (1974) - Identification in Control and Econometrics; Similarities and Differences. Annal of Economic and Social Measurement, 3/1.

MEHRA, R.K. (1979) - Kalman-Filters and their Applications to TIMS Studies in the Management Sciences, 12, pp. 15-94.

MIGON, H.S. and HARRISON, P.J. (1983) - An Application of Non-Linear Bayesian Forecasting to Television Advertising. Contributed paper at the 2nd Valencia International Meeting on Bayesian Statistics.

NELDER, J.A. (1968) - Weighted Regression, Quantal Response Data, and Inverse Polynomials. Biometrics, December 1968, pp. 979-985.

NELDER, J.A. and WEDDERBURN, R.W.M. (1972) - Generalized Linear Models. J.R. Statistic. Soc. A, 135, pp. 370-384.

NELSON, C.R. (1973) - Applied Time Series Analysis for Managerial Forecasting. San Francisco: Holden Day.

NEWBOLD, P. and GRANGER, C.W.J. (1974) - Experience with Forecasting Univariate Time series and the Combination of Forecasts. J.R. Statist. Soc. A., 137, Part 2, p. 131.

PAGAN, A. (1980) - Some Identification and Estimation Results for Regression Models with Stochastically varying Coefficients. Journal of Econometrics 13, pp. 341-363.

PIERCE, D.A. (1980) - A Survey on Recent Developments in Seasonal Adjustments. The American Statistician, Vol. 34, No. 3.

PRIESTLEY, M.B. (19) - Non-Linear Models in Time Series Analysis. The Statistician, Vol. 27, Nos. 3,4.

PRIESTLEY, M.B. (1980) - State-Dependent Models: A General Approach to Non-Linear Time Series Analysis. Journal of Time Series Analysis, Vol. 1, No. 1.

RAIFFA, H. and SCHLAIFFER, R. (1964) - Applied Statistical Decision Theory (3rd corrected printing). Harward: University Press. Boston.

ROBERTS, S.A. and HARRISON, P.J. (1983) - Parsimonious Modelling and Forecasting of Seasonal Time Series. Warwick Research Paper, 25.

SARRIS, A.H. (1973) - Kalman-Filter Models - a Bayesian Approach to Estimation of Time Varying Regression Coefficients. Annals of Economic and Social Measurement, 2/4, pp. 501-523.

SILVEY, S.D. (1975) - Statistical Inference. Chapman and Hall Ltd. London.

SIMON, J.L. and ARNDT, J. (1980) - The Shape of the Advertising Response Function. Journal of Advertising Research, Vol. 20, No. 4, pp. 11-27.

SMITH, J.Q. (1979) - A Generalization of the Bayesian Steady Forecasting Model. J.R. Statist. Soc. B, 41, No. 3, pp. 375-387.

SOUZA, R.C. (1981) - A Bayesian Entropy Approach to Forecasting: the Multi-State Model. Time Series Analysis (Houston-Texas). 535-542. North Holland.

STEVENS, C.F. (1977) - On the Variability of Demand for Family Items. Operational Research Quarterly, Vol. 25, No. 3.

STEWART, M.J. (1980) - Measuring Advertising Effects by area Tests. ADMAP, 110-12.

SWEETING, T. (1981) - Scale Parameters: A Bayesian Treatment. J.R. Statist. Soc., 43, No. 3, pp. 333-338.

TIAO, G.C. and BOX, G.E.P. (1981) - Modelling Multiple Time Series with Applications. Journal of the Americas Statistical Association, Vol. 76, No. 376, Theory and Methods Section.

TWYMAN, W.A. (1972) - Designing Advertising Research for Marketing Decisions. Journal of the Market Research Society, Vol. 15, No. 2.

WALLIS, K.F. (1982) - Seasonal Adjustment and Revision of Current Data: Linear Filters for the X-11 Method. J.R. Statist. Soc. A, 145, Part 1, pp. 74-85.

WEDDERBURN, R.W.M. (1974) - Quasi-Likelihood Functions, Generalized Linear Models, and the Grauss-Newton Method. Biometrika, 61, 3, 439-447.

WEDDERBURN, R.W.M. (1976) - On the Existence and Uniqueness of the Maximum Likelihood Estimates for Certain Generalized Linear Models. Biometrika, 63, 1, pp. 27-32.

WEST, M. (1983) - Generalized Linear Models Scale Parameters, Outlier Accommodation and Prior Distributions. 2nd Valencia International Meeting on Bayesian Statistics.

WEST, M., HARRISON, P.J. and MIGON, H.S. (198) - Dynamic Generalized Linear Models and Bayesian Forecasting. Warwick Research Report No. 41.

WILLIAMS, D.A. (1982) - Extra-Binomial Variation in Logistic Linear Models. Appl. Statist., 31, No. 2, pp. 144-148.